ONTAP et applications d’entreprise

Enterprise applications

NetApp
July 31, 2025

This PDF was generated from https://docs.netapp.com/fr-fr/ontap-apps-dbs/index.html on July 31, 2025.
Always check docs.netapp.com for the latest.

Sommaire

ONTAP et applications d'entreprise . . . 1
Epic . . . 2
Epic sur ONTAP . . . 2
Objectif . . . 2
Portée . . . 2
Public . . . 3
Epic sur ONTAP . . . 3
EPIC sur ONTAP . . . 3
Disponibilité Epic sur ONTAP . . . 4
Epic sur la consolidation ONTAP . . . 4
Epic sur l'efficacité ONTAP . . . 5
Performances Epic sur ONTAP . . . 5
Epic sur l'évolutivité ONTAP . . . 7
Configuration de l'efficacité du stockage Epic . . . 8
Epic sur la sécurité ONTAP . . . 8
Architecture et design Epic . . . 9
Architecture « Epic » . . . 9
Dimensionnement Epic . . . 12
Exigences de stockage Epic . . . 13
Architecture à quatre nœuds Epic . . . 14
Architecture Epic à 6 nœuds . . . 15
Architecture Epic à huit nœuds . . . 17
Configuration et meilleures pratiques . . . 18
Epic sur ONTAP - utilitaires hôtes . . . 18
Configuration de volume et de LUN Epic . . . 19
Protocoles Epic et de fichiers . . . 21
Gestion des performances Epic . . . 21
Epic sur ONTAP - protocoles . . . 21
Configuration de l'efficacité du stockage Epic . . . 22
Configuration de l'efficacité du stockage Epic . . . 22
Dimensionnement du stockage pour Epic . . . 24
Informations supplémentaires sur Epic sur ONTAP . . . 24
Documents d'orientation client Epic . . . 24
Hyper-V . . . 25
Instructions de déploiement et meilleures pratiques de stockage . . . 25
Présentation . . . 25
Stockage NetApp et environnement Windows Server . . . 26
Le provisionnement dans des environnements SAN . . . 30
Le provisionnement dans les environnements SMB . . . 39
Infrastructure de stockage Hyper-V sur NetApp . . . 42
Efficacité du stockage . . . 53
Sécurité . . . 56
Déploiement du serveur Nano . . . 56

Déployez le cluster Hyper-V. 60
Déployer Hyper-V Live migration dans un environnement en cluster 61
Déployer Hyper-V Live migration en dehors d'un environnement en cluster 62
Déployez Hyper-V Storage Live migration 63
Déployer le réplica Hyper-V en dehors d'un environnement en cluster 64
Déployer la réplique Hyper-V dans un environnement en cluster 65
Où trouver des informations complémentaires 66
Microsoft SQL Server 68
Présentation 68
Workloads Microsoft SQL Server 68
Configuration de la base de données 69
Configuration du processeur 69
Configuration de la mémoire 72
Instance partagée ou instance dédiée 75
Fichiers tempdb 76
Configuration du stockage sur les systèmes AFF/FAS 77
Présentation 77
Fichiers de base de données et groupes de fichiers 80
Efficacité du stockage 84
Protection des données 89
Reprise après incident 92
Configuration du stockage sur les systèmes ASA r2 113
Présentation 113
Fichiers de base de données et groupes de fichiers 115
Protection des données 120
Reprise après incident 121
Sécurité des données 137
Copies Snapshot 137
Des snapshots inviolables 137
Réplication SnapMirror 138
Ordinateurs virtuels de stockage 138
RBAC d'administration 138
Authentification multifacteur (MFA) 138
RBAC D'API 139
Vérification multiadministrateur 139
MySQL 140
Présentation 140
Configuration de la base de données 140
Structure de fichier 140
Paramètres de configuration 143
taille_fichier_log_innodb 144
innodb_flush_log_at_trx_commit 144
innodb_doublewrite 145
innodb_buffer_pool_size 145
innodb_flush_method 146

innodb_io_capacity . . . 146
innodb_lru_scan_depth . . . 147
open_file_limits . . . 147
Configuration de l'hôte . . . 148
Conteneurisation . . . 148
Tables d'emplacements NFSv3 . . . 148
Planificateurs d'E/S. . . . 149
Descripteurs de fichier . . . 149
Configuration de stockage sous-jacente . . . 150
NFS . . . 150
SAN . . . 151
Base de données Oracle . . . 153
Bases de données Oracle sur ONTAP . . . 153
Configuration ONTAP . . . 153
RAID . . . 153
Gestion de la capacité . . . 154
Ordinateurs virtuels de stockage . . . 155
Gestion des performances grâce à la QoS de ONTAP . . . 155
Efficacité . . . 157
Provisionnement fin . . . 161
Basculement/basculement ONTAP . . . 163
Configuration de la base de données . . . 165
Tailles de bloc . . . 165
db_file_multibloc_read_count . . . 166
filesytemio_options . . . 167
Délais d'expiration du RAC . . . 168
Configuration de l'hôte . . . 169
AIX . . . 169
HP-UX . . . 171
Linux . . . 173
ASMLib/AFD (pilote de filtre ASM) . . . 177
Microsoft Windows . . . 179
Solaris . . . 179
Configuration du réseau . . . 185
Interfaces logiques . . . 185
Configuration TCP/IP et ethernet . . . 189
Configuration FC SAN . . . 191
Réseau à connexion directe . . . 192
Configuration de stockage sous-jacente . . . 192
SAN FC . . . 193
NFS . . . 197
NVFAIL . . . 210
Utilitaire de récupération ASM (ASMRU) . . . 211
Virtualisation . . . 211
Prise en charge . . . 211

Présentation du stockage . . . 211
Pilotes paravirtualisés . . . 213
Saturation de la mémoire RAM . . . 213
Répartition des datastores . . . 213
Tiering . . . 214
Présentation . . . 214
Règles de hiérarchisation . . . 216
Stratégies de Tiering . . . 218
Interruptions d'accès au magasin d'objets . . . 222
Protection des données Oracle . . . 222
Protection des données avec ONTAP . . . 222
Planification des objectifs de durée de restauration, de point de récupération et des SLA . . . 223
Disponibilité de la base de données . . . 226
Checksums et intégrité des données . . . 228
Notions de base sur la sauvegarde et la restauration . . . 233
Reprise sur incident Oracle . . . 247
Présentation . . . 247
MetroCluster . . . 248
Synchronisation active SnapMirror . . . 268
Migration de la base de données Oracle . . . 302
Présentation . . . 302
Planification des migrations . . . 303
Procédures . . . 306
Exemples de scripts . . . 413
Remarques supplémentaires . . . 426
Optimisation des performances et analyse comparative . . . 426
Les verrous NFSv3 obsolètes . . . 429
Vérification de l'alignement WAFL . . . 430
PostgreSQL . . . 436
Présentation . . . 436
Configuration de la base de données . . . 436
Architecture . . . 436
Paramètres d'initialisation . . . 437
Paramètres . . . 438
Espaces de stockage . . . 440
Configuration de stockage sous-jacente . . . 440
NFS . . . 440
SAN . . . 442
Protection des données . . . 444
Protection Ddta native . . . 444
Snapshots . . . 445
Logiciels de protection des données . . . 446
SAP . . . 447
VMware . . . 448
VMware vSphere avec ONTAP . . . 448

VMware vSphere avec ONTAP . . . 448
Pourquoi choisir ONTAP pour VMware vSphere ? . . . 448
Stockage unifié . . . 450
Outils de virtualisation pour ONTAP . . . 452
Volumes virtuels (vvols) et gestion basée sur des règles de stockage (SPBM) . . . 454
Datastores et protocoles . . . 455
Configuration du réseau . . . 470
Clonage des VM et des datastores . . . 473
Protection des données . . . 475
La qualité de service (QoS) . . . 478
Migration et sauvegarde dans le cloud . . . 483
Chiffrement pour les données vSphere . . . 484
Active IQ Unified Manager . . . 485
Gestion basée sur des règles de stockage et vVols . . . 487
Planificateur de ressources distribué de stockage VMware . . . 489
Hôte ESXi recommandé et autres paramètres ONTAP recommandés . . . 490
Volumes virtuels (vVols) avec les outils ONTAP 10 . . . 494
Présentation . . . 494
Liste de contrôle . . . 500
Utilisation de vVols avec ONTAP . . . 502
Déploiement de vVols sur les systèmes AFF, ASA, ASA r2 et FAS . . . 508
Protection des vVols . . . 519
Dépannage . . . 524
VMware site Recovery Manager et ONTAP . . . 525
Restauration de site en direct VMware avec ONTAP . . . 525
Bonnes pratiques de déploiement . . . 527
Meilleures pratiques opérationnelles . . . 528
Topologies de réplication . . . 533
Dépannage de VLSRM/SRM lors de l'utilisation de la réplication vVols . . . 541
Informations supplémentaires . . . 542
Cluster de stockage vSphere Metro avec ONTAP . . . 542
Cluster de stockage vSphere Metro avec ONTAP . . . 542
Présentation de la solution VMware vSphere . . . 545
Directives de conception et de mise en œuvre VMSC . . . 550
Résilience pour les événements planifiés et non planifiés . . . 561
Scénarios de défaillance pour vMSC avec MetroCluster . . . 562
Sécurité des produits . . . 573
Les outils ONTAP pour VMware vSphere . . . 573
Plug-in SnapCenter VMware vSphere . . . 575
Guide de renforcement de la sécurité des outils ONTAP pour VMware vSphere . . . 577
Guide de renforcement de la sécurité des outils ONTAP pour VMware vSphere 9.13 . . . 578
Vérification de l'intégrité des outils ONTAP pour les packages d'installation de VMware vSphere 9.13 . . . 578
Ports et protocoles pour les outils ONTAP 9.13 . . . 580
Outils ONTAP pour les points d'accès VMware vSphere 9.13 (utilisateurs) . . . 581
ONTAP Tools 9.13 Mutual TLS (authentification basée sur certificat) . . . 582

Certificat HTTPS des outils ONTAP 9.13 . . . 588
Bannière de connexion Outils ONTAP 9.13 . . . 588
Délai d'inactivité pour les outils ONTAP 9.13 . . . 589
Nombre maximal de requêtes simultanées par utilisateur (protection de la sécurité réseau/attaque dos) Outils ONTAP pour VMware vSphere 9.13 . . . 589
Configuration du protocole NTP (Network Time Protocol) pour les outils ONTAP 9.13 . . . 590
Stratégies de mot de passe pour les outils ONTAP 9.13 . . . 590
Mentions légales . . . 592
Droits d'auteur . . . 592
Marques déposées . . . 592
Brevets . . . 592
Politique de confidentialité . . . 592
Source ouverte . . . 592
ONTAP . . . 592
ONTAP Mediator pour les configurations IP MetroCluster . . . 592

ONTAP et applications d’entreprise

Epic

Epic sur ONTAP

La clé de la transformation digitale consiste à en faire plus avec vos données.

> Cette documentation remplace le rapport technique *TR-3923 : meilleures pratiques de NetApp pour Epic*.

Les hôpitaux exigent d'importants volumes de données pour amorcer la transformation digitale. Une partie du processus de traitement des patients, de gestion des horaires du personnel et des ressources médicales consiste à recueillir et à traiter les informations. Cependant, de nombreuses actions sont toujours exécutées manuellement ou via des systèmes obsolètes. La seule constante est que la quantité de données continue à croître de manière exponentielle et donc, devient de plus en plus difficile à gérer.

La principale cause de ce problème est que les données hospitalières sont souvent stockées dans des silos de données. Trop de temps est consacré aux entrées et mises à jour manuelles, ce qui entraîne des pannes et des erreurs. Ce document s'applique à environ une partie des données de santé, soit le dossier médical électronique (DME) Epic. Cependant, la stratégie de gestion des données couverte ici peut et doit être appliquée à toutes les données de santé. Depuis longtemps déjà, NetApp modernise et simplifie l'infrastructure digitale. Notre infrastructure de données intelligente est à la base de la transformation digitale.

NetApp propose une solution unique de gestion des données pour tous les besoins de santé. Nous pouvons guider les hôpitaux tout au long de leur transition vers la transformation digitale. En construisant une base avec une structure et des solutions intelligentes, les soins de santé peuvent extraire toute la valeur de ces précieuses informations. Ce cadre peut aider les médecins à diagnostiquer les maladies plus rapidement et à élaborer des plans de traitement individualisés pour mieux soutenir les processus de prise de décision dans les situations d'urgence. Vous pourrez également créer votre propre infrastructure de données intelligente et permettre à votre hôpital d'exploiter les silos de données, de faciliter l'interopérabilité des données et de protéger les informations confidentielles des patients.

Ce document vous aidera à construire et à mettre en œuvre avec succès des DME EPIC. Plutôt que de créer plusieurs silos Epic, créez une infrastructure de données Epic unique et transformez votre hôpital.

Objectif

Ce document présente les bonnes pratiques d'intégration du stockage NetApp dans un environnement logiciel Epic. Il contient les sections suivantes :

- Connaissance technique de l'environnement logiciel Epic et de ses besoins en stockage dans différentes configurations.
- Considérations relatives au stockage Epic et description des facteurs importants de prise de décision pour les solutions Epic.
- Recommandations sur le stockage NetApp, décrivant les bonnes pratiques de configuration du stockage NetApp et permettant de répondre aux besoins de stockage d'Epic.

Portée

Ce document ne couvre pas les sujets suivants :

- Exigences quantitatives en matière de performances et conseils de dimensionnement, qui sont abordés

dans le "Tr-3930i : instructions de dimensionnement NetApp pour Epic" (connexion NetApp requise)

Public

NetApp suppose que le lecteur possède les connaissances de base suivantes :

- Une solide compréhension des concepts SAN et NAS
- Connaissance technique des systèmes de stockage ONTAP
- Connaissance technique de la configuration et de l'administration de ONTAP

Epic sur ONTAP

EPIC sur ONTAP

Epic est plus facile avec ONTAP.

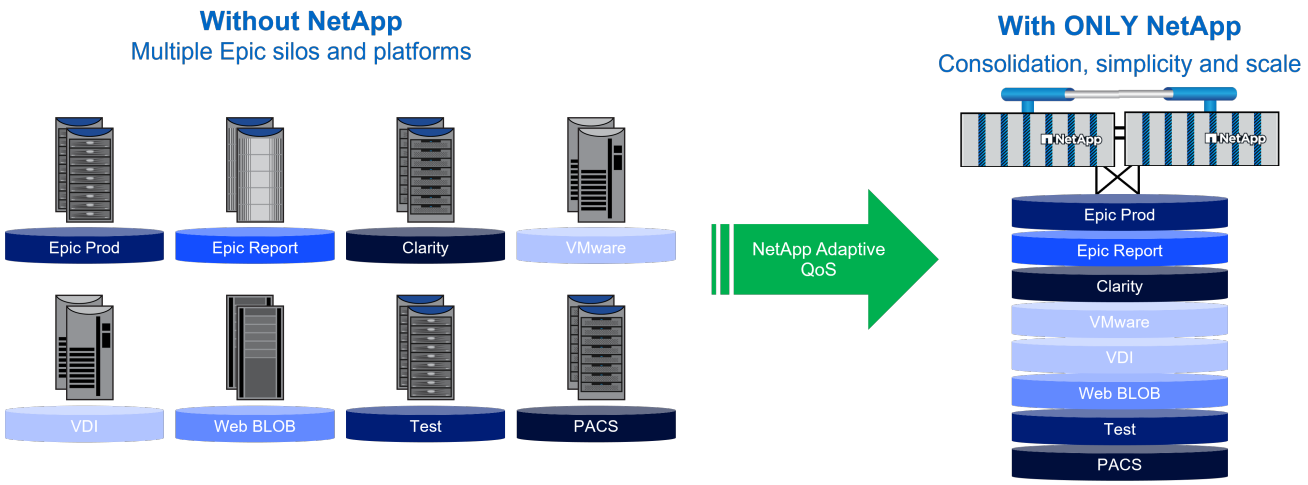
ONTAP est une plateforme de gestion des données qui vous permet de consolider les workloads Epic tout en répondant à vos besoins de performances, de protection des données et de gestion des données.

Seul NetApp permet de standardiser tous vos workloads de santé pour SAN, NAS et objet sur une plateforme unique de gestion des données haute disponibilité. ONTAP est la plateforme logicielle de stockage la plus largement déployée au monde et s'accompagne de près de 30 ans d'innovation constante. Vous pouvez relever tous vos défis Epic grâce aux outils de gestion des données ONTAP natifs et à l'intégration des applications. Nul besoin d'acheter une multitude d'outils tiers pour combler les lacunes de la solution.

De nombreux fournisseurs de stockage proposent un stockage bloc classique, fiable et rapide. Elles fonctionnent bien, mais sont généralement déployées en silos pour exécuter une seule charge de travail, telle que la production, le reporting, la clarté, l'infrastructure de postes de travail virtuels, VMware et NAS. Chacun de ces silos possède du matériel et des outils de gestion différents, et sont généralement gérés par des groupes INFORMATIQUES différents. Cette approche traditionnelle ajoute au plus gros problème dans le domaine de la santé aujourd'hui : la complexité.

NetApp facilite et optimise la gestion des données. Au lieu d'investir dans des silos surdimensionnés, ONTAP utilise l'innovation et la technologie pour fournir un niveau de service cohérent et garanti pour chaque charge de travail sur une plateforme unique, quel que soit le protocole utilisé avec la protection des données intégrée. Ces fonctionnalités et outils s'étendent également au cloud de votre choix, comme illustré ci-dessous.

Scale and simplicity for Healthcare



Disponibilité Epic sur ONTAP

La continuité de l'activité est au cœur de ONTAP et vous permet d'éviter les interruptions coûteuses de vos activités.

NetApp assure une disponibilité supérieure à 99.999999 % sur les données de production, qui sont appelées « maison » via NetApp Active IQ. Chaque paire haute disponibilité dans le cluster ne présente pas de point de défaillance unique. Depuis 1992, ONTAP est le logiciel de gestion des données le plus largement déployé au monde. Depuis, sa fiabilité de stockage a fait ses preuve d'une expérience exceptionnelle. Active IQ surveille et résout automatiquement 97 % des problèmes de manière proactive. La disponibilité est ainsi accrue et le nombre de dossiers de demande de support est considérablement réduit.

Epic recommande l'utilisation de systèmes de stockage haute disponibilité pour limiter les défaillances de composants matériels. Cette recommandation s'étend du matériel de base (par exemple, les alimentations redondantes) aux réseaux (par exemple, les réseaux à chemins d'accès multiples).

Lorsque vous devez mettre à niveau le stockage, le faire évoluer verticalement, le faire évoluer horizontalement ou rééquilibrer les charges de travail dans le cluster, les soins aux patients ne sont pas pris en charge. Vous pouvez déplacer des données, mais les patients ne doivent plus jamais interrompre les soins lors des migrations de données ou des mises à niveau majeures. Optez pour une technologie nouvelle génération, pérenne et sans dépendance vis-à-vis d'un seul matériel. NetApp garantit même une disponibilité écrite de 100 %.

Pour plus d'informations sur la fiabilité, la disponibilité, la facilité de maintenance et les fonctions de sécurité de NetApp, consultez le "Fiabilité, disponibilité, facilité de maintenance et sécurité du système NetApp ONTAP"livre blanc.

Epic sur la consolidation ONTAP

L'un des principaux défis dans le domaine de la santé est l'inefficacité des environnements en silos.

Plusieurs solutions ponctuelles sont créées par divers groupes qui empêchent le progrès. Une stratégie unifiée de gestion des données apporte de l'efficacité pour accélérer la transformation. Des technologies

perturbatrices, telles que la numérisation des dossiers médicaux, les ransomwares et l'IA générative, qui sont autant d'éléments qui nécessitent une consolidation.

Avec ONTAP, vous pouvez consolider les fichiers, les blocs et les objets ainsi que chacun de vos workloads de Tier 0/1/2/3, sur site et dans le cloud, qui s'exécutent tous sur ONTAP.

Epic sur l'efficacité ONTAP

Epic s'exécute sur des baies 100 % Flash, où la majeure partie du coût est le disque. Par conséquent, l'efficacité du stockage est essentielle pour réaliser des économies.

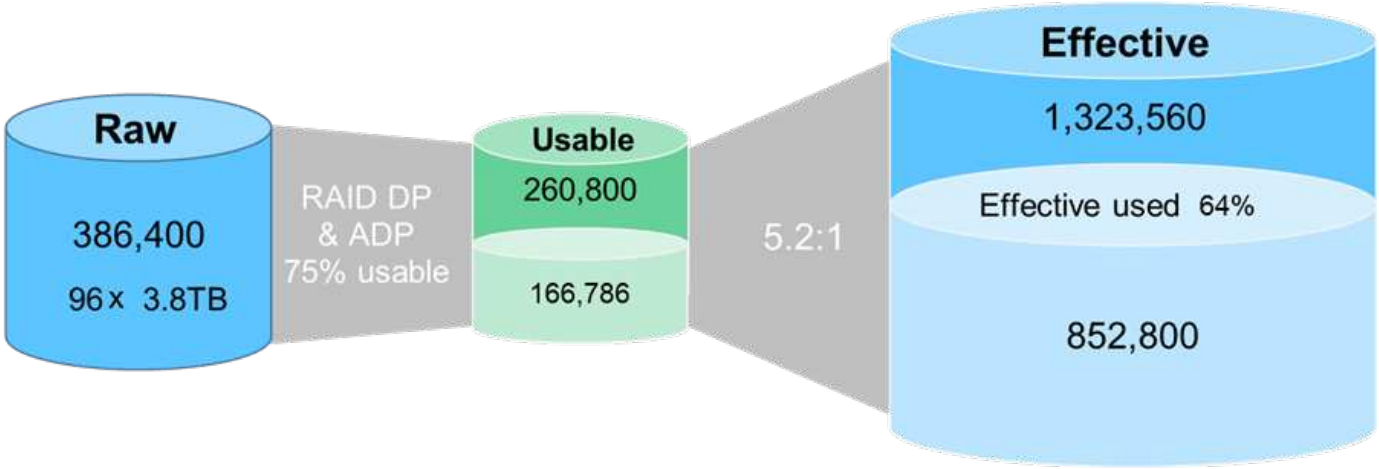
L'efficacité du stockage à la volée de NetApp permet de réaliser des économies de stockage de pointe sans effets sur les performances, et nous proposons même une garantie d'efficacité écrite avec les baies 100 % Flash.

Lors du calcul de l'efficacité du stockage, il est important de mesurer la capacité brute à la capacité utilisable réelle.

- **Capacité brute** avant d'appliquer un RAID, taille du disque par nombre de disques.
- **Capacité utilisable** après application de RAID, combien de stockage utilisable est disponible.
- **Capacité effective** quantité de stockage provisionnée et présentée à l'hôte ou au client.

La figure ci-dessous est un exemple de calcul d'efficacité pour un déploiement Epic classique, incluant toutes les charges de travail nécessitant 852 To de stockage effectif et avec un ratio d'efficacité de 5.2:1 pour 1,32 po de données effectives totales.

La capacité brute à utilisable varie légèrement en fonction du nombre de disques.



NetApp n'utilise pas la technologie Snapshot de NetApp ni le provisionnement fin pour calculer l'efficacité du programme de garantie. Cela démontre une efficacité irréaliste entre 30 et 100:1, ce qui ne veut rien dire lors du dimensionnement de la capacité de stockage réelle.

Performances Epic sur ONTAP

ONTAP a introduit les technologies Flash en 2009 et prend en charge les disques SSD depuis 2010. Cette longue expérience du stockage Flash permet à NetApp d'ajuster les fonctionnalités de ONTAP afin d'optimiser les performances des disques SSD et

d’améliorer la longévité des supports Flash tout en conservant les fonctionnalités avancées de ONTAP.

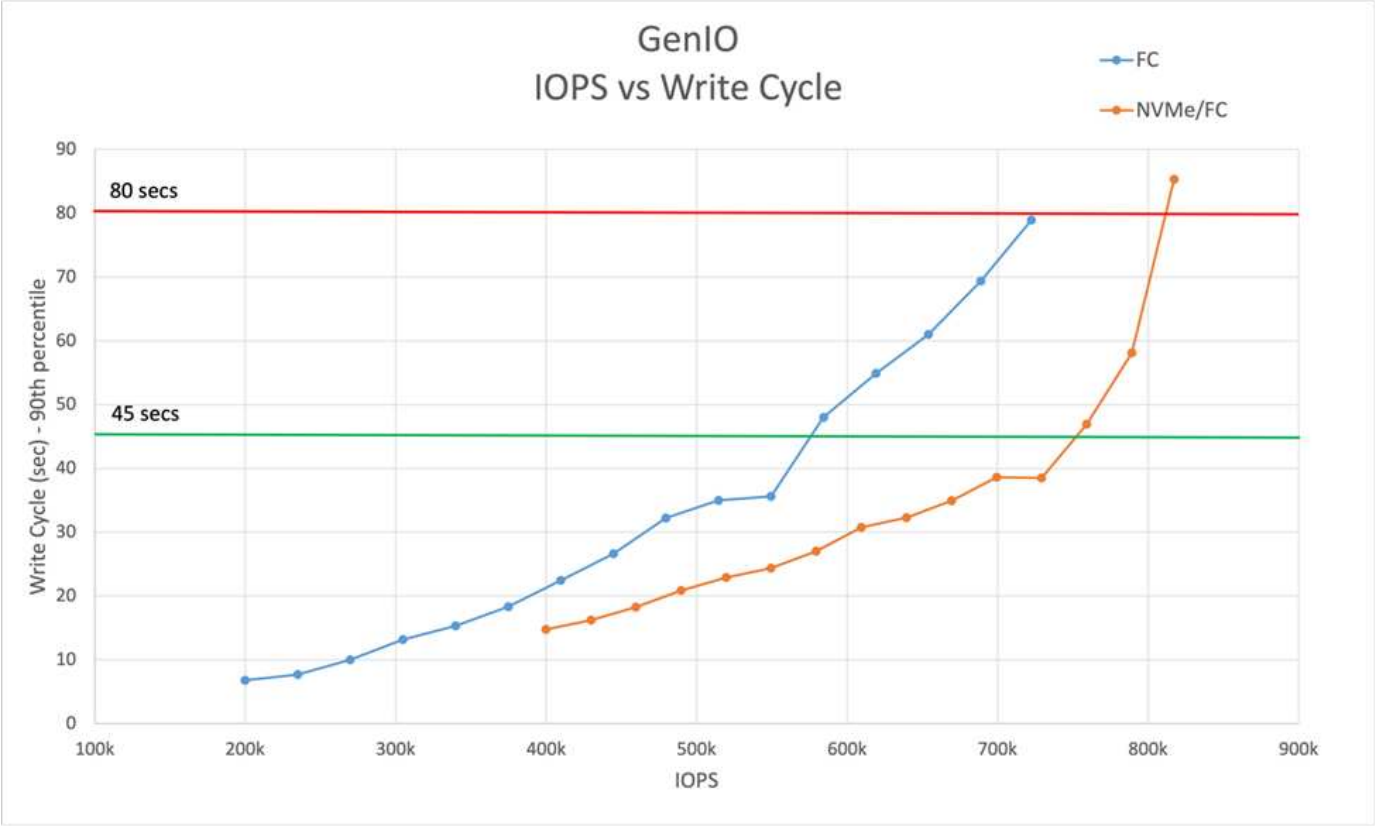
Depuis l’année 2020, tous les workloads ODB d’Epic doivent être hébergés sur du stockage 100 % Flash. Les charges de travail Epic fonctionnent généralement à environ 1,000 à 2,000 IOPS par téraoctet de stockage (bloc de 8 Ko, ratio lecture/écriture de 75 à 25 % et aléatoire de 100 %). Epic est très sensible à la latence, et une latence élevée a un impact visible sur l’expérience utilisateur, ainsi que sur les tâches opérationnelles telles que l’exécution de rapports, de sauvegardes, de vérifications d’intégrité et de mises à jour de l’environnement.

- Les baies 100 % Flash n’étant pas le facteur limitant, elles se limitent aux disques, mais à l’utilisation des contrôleurs.
- ONTAP utilise une architecture actif-actif. Pour les performances, les deux nœuds de la paire haute disponibilité écrivent sur les disques.
- Il en résulte une utilisation optimisée du processeur, qui est le facteur le plus important qui permet à NetApp de publier les meilleures performances Epic du secteur.
- Les technologies NetApp RAID DP, ADP (Advanced Disk Partitioning) et WAFL répondent à tous les besoins Epic. Toutes les charges de travail distribuent les E/S sur tous les disques. Sans goulot d’étranglement.
- ONTAP est optimisé pour l’écriture. Les écritures sont reconnues une fois écrites sur la NVRAM en miroir avant d’être écrites sur le disque à la vitesse de la mémoire en ligne.
- WAFL, NVRAM et l’architecture modulaire permettent à NetApp d’utiliser le logiciel pour innover avec des fonctionnalités d’efficacité à la volée, de chiffrement et de performance. Ils permettent également à NetApp de lancer de nouvelles fonctionnalités sans affecter les performances.
- Jusqu’à présent, chaque nouvelle version de ONTAP a enregistré une augmentation des performances et de l’efficacité de l’ordre de 30 à 50 %. Toujours à jour avec ONTAP, les performances sont optimales.

NVMe

Lorsque les performances sont primordiales, NetApp prend également en charge NVMe/FC, le protocole FC SAN nouvelle génération.

Comme le montre la figure ci-dessous, nos tests Genio ont réalisé un nombre beaucoup plus élevé d’IOPS en utilisant le protocole NVMe/FC par rapport au protocole FC. La solution connectée NVMe/FC a atteint plus de 700 000 IOPS avant de dépasser le seuil du cycle d’écriture de 45 secondes. Le remplacement des commandes SCSI par NVMe permet également de réduire considérablement l’utilisation sur l’hôte.



Epic sur l'évolutivité ONTAP

Le Guide de configuration du matériel Epic représente environ 20 % de croissance annuelle pendant 3 ans. Cependant, les environnements peuvent également se développer de manière inattendue.

NetApp peut facilement faire évoluer les performances et la capacité jusqu'à 12 nœuds pour les clusters NAS, SAN et objet. Par conséquent, vous pouvez évoluer horizontalement et verticalement sans interruption pour accompagner la croissance de votre entreprise.

EPIC Iris offre des capacités d'évolutivité supplémentaires. Il permet aux clients de plus grande taille disposant de plusieurs instances Epic de procéder à la consolidation vers une seule instance. "Epic sur l'architecture vérifiée NetApp sur le SAN moderne"Ce document démontre qu'Epic peut faire évoluer de manière fluide les charges de travail consolidées à 720 000 IOPS sur une seule haute disponibilité et évoluer en mode scale-out jusqu'à plus de 4 millions d'IOPS dans un cluster. Vous pouvez évoluer en mode « scale-up » sans interrompre l'activité en mettant à niveau les contrôleurs ou en ajoutant des disques aux clusters existants.

Les données NAS, SAN et objet peuvent également être déplacées sans interruption entre les nœuds du cluster. Chaque paire haute disponibilité dans le cluster peut correspondre à n'importe quelle combinaison de types et de tailles de systèmes ONTAP FAS et AFF. Vous pouvez équilibrer vos charges de travail sur un seul cluster afin d'optimiser votre investissement en stockage.

ONTAP permet également d'utiliser le stockage objet sur StorageGRID ou le cloud comme cible de sauvegarde et/ou cible de Tiering automatique du stockage à froid. Ainsi, vous pouvez libérer automatiquement des disques 100 % Flash coûteux, des snapshots de Tiering et des données inactives dans un environnement objet.

Le résultat : Epic fonctionne simplement mieux avec le portefeuille de produits NetApp, exploitant ONTAP, plusieurs protocoles, StorageGRID et le cloud de votre choix. Ces produits proposent des options de reprise

après incident, d'archivage, d'analytique, de Tiering, etc.

Configuration de l'efficacité du stockage Epic

Un snapshot est une copie instantanée d'un volume en lecture seule.

Un snapshot place un verrou logique sur tous les blocs du système de fichiers actif. Les copies Snapshot NetApp ONTAP sont quasi instantanées et n'utilisent aucun stockage supplémentaire.

Write anwhere File Layout, ou WAFL, est un système de fichiers en écriture seule ; il n'effectue pas d'E/S supplémentaires, telles que la copie des données dans un bloc protégé par copie Snapshot avant d'être écrasé. Aucune donnée n'est jamais déplacée. Par conséquent, les copies Snapshot n'ont aucun impact sur la capacité de stockage ou les performances. Les copies Snapshot permettent de réaliser des économies de stockage considérables tout en augmentant la solution de sauvegarde.

FlexClone

Un volume NetApp ONTAP FlexClone est un clone d'un volume existant ou un snapshot d'un volume existant. Il s'agit autrement d'un volume ONTAP comme tout autre, et il peut lui-même être cloné, protégé par des copies Snapshot et configuré avec une règle de qualité de service.

Comme pour les snapshots, un volume FlexClone ne nécessite aucun espace supplémentaire au moment de sa création. Seules les modifications apportées au clone nécessitent une capacité supplémentaire.

Epic requiert entre 10 et 30 copies des bases de données de production pour répondre à divers besoins opérationnels, tels que les sauvegardes en continu, les vérifications d'intégrité et les environnements de mise à niveau intermédiaire. La nécessité d'une solution basée sur les volumes FlexClone s'est accrue avec le passage à des mises à niveau plus fréquentes.

> Une solution de sauvegarde Epic entièrement automatisée et une solution d'actualisation Epic sont fournies par NetApp dans le cadre de la solution à l'aide d'Ansible et des outils NetApp natifs.

Epic sur la sécurité ONTAP

La sécurité est aujourd'hui la principale préoccupation des entreprises et des cadres de la santé. Sa gestion n'a jamais été aussi difficile et les entreprises sont confrontées à des défis tels que la conformité, la gouvernance des données, la protection antivirus et les ransomwares.

Le présent document ne contient pas de guide complet sur Epic et la sécurité du stockage. Il détaille toutefois "Guide sur le renforcement de la sécurité de la solution ONTAP"toutes les fonctionnalités de sécurité avancées et étendues disponibles avec ONTAP.

NetApp Active IQ Unified Manager surveille les violations de sécurité en fonction des informations incluses dans et les signale dans le "TR-4569"tableau de bord afin de simplifier la gestion de la sécurité. Ces outils peuvent aider votre entreprise à atteindre ses objectifs de sécurité en matière de protection, de détection et de résolution des attaques.

NetApp a également établi des partenariats avec des fournisseurs de solutions de sécurité pour permettre une intégration par le biais de "NetApp FPolicy" logiciels permettant d'améliorer vos offres de sécurité. En outre, "Authentification multifacteur (MFA)" ils peuvent être ajoutés pour sécuriser votre environnement Epic contre tout accès non autorisé avec des informations d'identification qui ont fui.

Enfin, les copies Snapshot natives ONTAP et les technologies SnapLock immuables avec "Cyber-coffre ONTAP", offrent une fonctionnalité d'air Gap unique pour protéger vos dossiers médicaux contre les ransomwares. Voir la documentation NetApp sur "Solution NetApp pour ransomware". Pour une approche plus stratégique de la sécurité, voir "NetApp et le modèle « zéro confiance »".

Architecture et design Epic

Architecture « Epic »

Cette section décrit l'environnement logiciel Epic et les composants clés qui nécessitent un stockage. Fournit les principaux éléments à prendre en compte pour guider la conception du stockage.

Epic, dont le siège social se trouve à Vérone, dans le Wisconsin, propose des logiciels pour les groupes médicaux de taille moyenne à grande, les hôpitaux et les organismes de santé intégrés. Les clients comprennent également des hôpitaux communautaires, des établissements universitaires, des organisations pour enfants, des fournisseurs de filet de sécurité et des systèmes multi-hospitaliers. Le logiciel intégré au système Epic couvre les fonctions cliniques, d'accès et de revenu et s'étend à la maison.

Il n'est pas question du champ d'application de ce document de couvrir le large éventail de fonctions prises en charge par le logiciel Epic. Cependant, pour le système de stockage, tous les logiciels Epic partagent une base de données unique axée sur le patient pour chaque déploiement. EPIC passe de la base de données InterSystems Caché à la nouvelle base de données InterSystems Iris. Comme les exigences de stockage sont les mêmes pour Caché et Iris, nous nous référons à la base de données comme Iris tout au long du reste de ce document. Iris est disponible pour les systèmes d'exploitation AIX et Linux.

Iris InterSystems

InterSystems Iris est la base de données utilisée par l'application Epic. Dans cette base de données, le serveur de données est le point d'accès des données stockées de manière persistante. Le serveur d'applications gère les requêtes de base de données et envoie des requêtes de données au serveur de données. Pour la plupart des environnements logiciels Epic, l'utilisation de l'architecture multiprocesseur symétrique (SMP) dans un seul serveur de base de données suffit pour répondre aux demandes de base de données des applications Epic. Dans les déploiements de grande envergure, un modèle distribué peut être pris en charge à l'aide du protocole ECP (Enterprise Caché Protocol) d'InterSystems.

L'utilisation d'un matériel en cluster prenant en charge le basculement permet à un serveur de données de secours d'accéder au même stockage que le serveur de données principal. Il permet également au serveur de données de secours de prendre en charge les responsabilités de traitement en cas de panne matérielle.

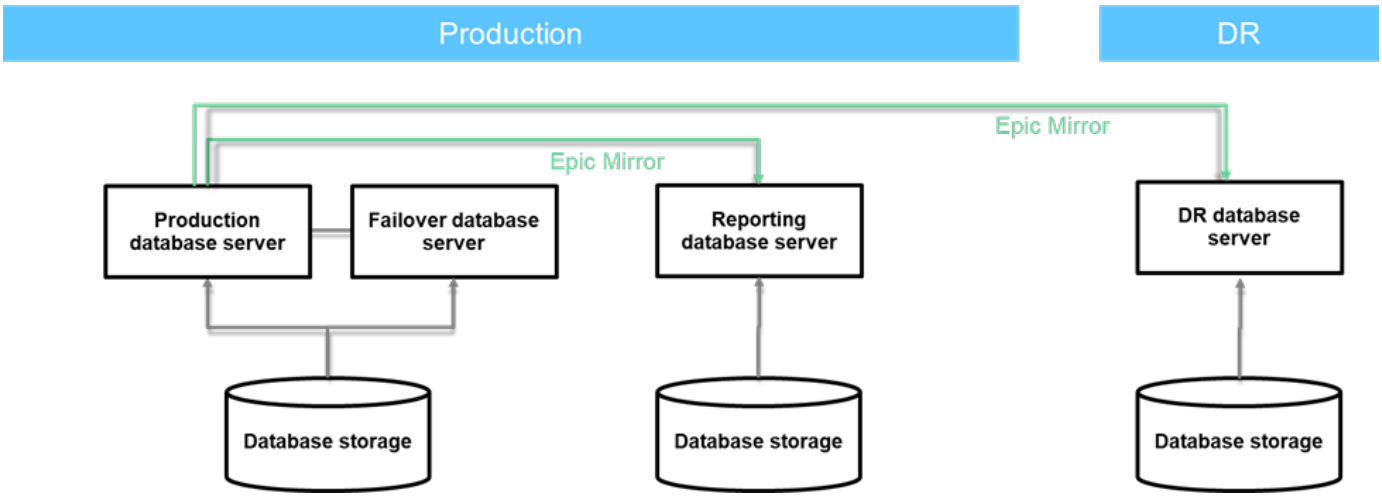
InterSystems fournit également des technologies qui répondent aux besoins en termes de réplication des données, de reprise après incident et de haute disponibilité. La technologie de réplication d'InterSystems permet de répliquer une base de données Iris de manière synchrone ou asynchrone depuis un serveur de données principal vers un ou plusieurs serveurs de données secondaires. NetApp SnapMirror est utilisé pour répliquer le stockage WebBLOB ou pour la sauvegarde et la reprise après incident.

La mise à jour de la base de données Iris présente de nombreux avantages :

- Évolutivité accrue et consolidation en une plus grande instance grâce à plusieurs instances Epic.
- Offre de licence permettant aux clients de passer d'AIX à Red Hat Enterprise Linux (RHEL) sans payer de nouvelle licence de plate-forme.

Serveurs de base de données Caché et utilisation du stockage

- **Production** dans les environnements logiciels Epic, une base de données unique centrée sur le patient est déployée. Dans les exigences matérielles d'Epic, le serveur physique hébergeant le serveur de données Iris de lecture/écriture principal est appelé serveur de base de données de production. Ce serveur nécessite un stockage 100 % Flash haute performance pour les fichiers appartenant à l'instance de base de données primaire. Pour la haute disponibilité, Epic prend en charge l'utilisation d'un serveur de base de données de basculement ayant accès aux mêmes fichiers. Iris utilise Epic Mirror pour répliquer en lecture seule des rapports, la reprise après incident et la prise en charge des copies en lecture seule. Chaque type de serveur de base de données peut être basculé en mode lecture/écriture pour des raisons de continuité d'activité.
- **Report** Un serveur de base de données miroir fournit un accès en lecture seule aux données de production. Il héberge un serveur de données Iris configuré comme miroir de sauvegarde du serveur de données Iris de production. Les besoins en capacité de stockage du serveur de la base de données de production sont les mêmes que ceux du serveur de la base de données de production. Les rapports de performances d'écriture sont identiques à ceux de la production, mais les caractéristiques de la charge de travail de lecture sont différentes et dimensionnées différemment.
- **Prend en charge la lecture seule** ce serveur de base de données est facultatif et n'est pas illustré ci-dessous. Un serveur de base de données en miroir peut également être déployé pour prendre en charge la fonctionnalité Epic en lecture seule, dans laquelle l'accès est fourni à une copie de production en mode lecture seule. Ce type de serveur de base de données peut être basculé en mode lecture/écriture pour des raisons de continuité d'activité.
- **Récupération après sinistre** pour atteindre les objectifs de continuité de l'activité et de reprise après sinistre, un serveur de base de données miroir de reprise après sinistre est généralement déployé sur un site distinct géographiquement des serveurs de base de données miroir de production et/ou de génération de rapports. Un serveur de base de données miroir de reprise après sinistre héberge également un serveur de données Iris configuré comme miroir de sauvegarde du serveur de données Iris de production. Si le site de production devient indisponible pendant une période prolongée, ce serveur de base de données miroir de sauvegarde peut être configuré pour agir en tant qu'instance de lecture/écriture miroir (SRW). Le serveur de base de données miroir de sauvegarde a les mêmes besoins en stockage de fichiers que le serveur de base de données de production. En revanche, le stockage de la base de données en miroir de sauvegarde est dimensionné de la même manière que le stockage de production du point de vue des performances, pour assurer la continuité de l'activité.



- **Test** les organismes de santé déploient souvent des environnements de développement, de test et de transfert. Les serveurs de données Iris supplémentaires pour ces environnements nécessitent également un stockage, qui peut être pris en charge par le même système de stockage. Epic présente des exigences

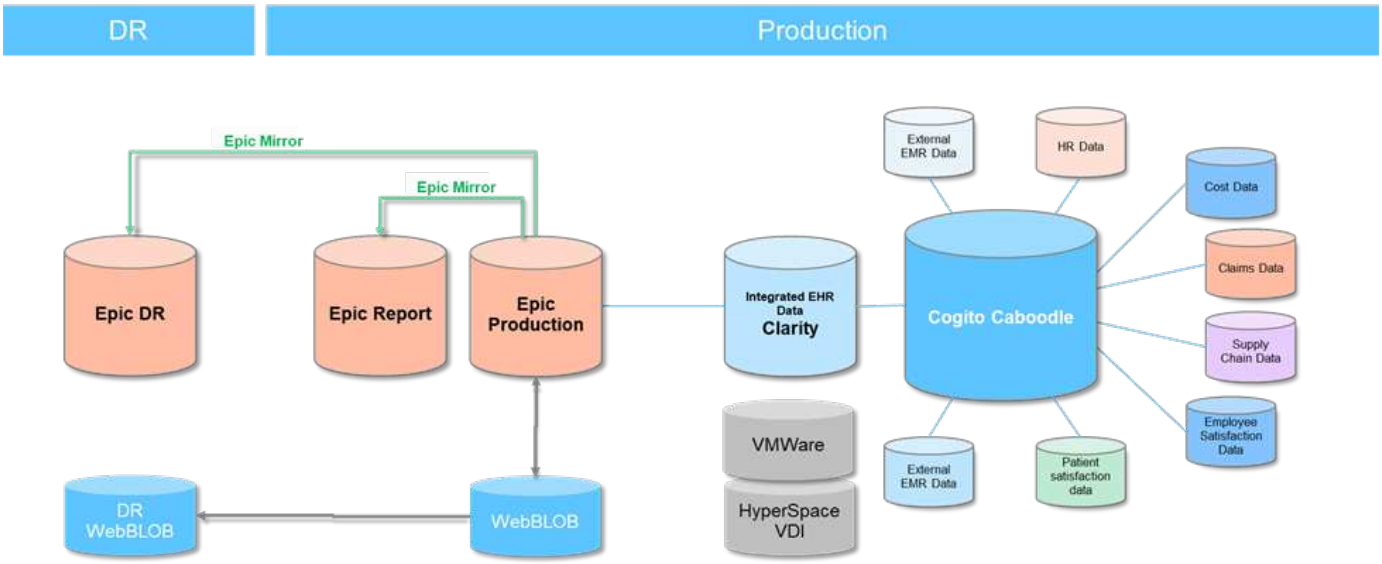
et des contraintes spécifiques pour fournir du stockage supplémentaire à partir d'un système de stockage partagé. Ces exigences spécifiques sont traitées de façon générique par les meilleures pratiques de ce document.

Outre les serveurs de données ODB Iris, les environnements logiciels Epic incluent généralement d'autres composants, tels que les suivants et comme illustré dans la figure ci-dessous :

- Serveur de base de données Oracle ou Microsoft SQL Server en tant que back-end des outils de reporting d'entreprise Clarity d'Epic

> Clarity est utilisé pour générer des rapports sur les données extraites quotidiennement de la base de données Iris.

- Serveur WebBLOB (SMB)
- Serveur de base de données polyvalent
- Machines virtuelles polyvalentes
- Hyperspace pour l'accès client



Les besoins en stockage de toutes ces charges de travail multiples, pools, protocoles NAS et SAN peuvent être consolidés et hébergés par un seul cluster ONTAP. Cette consolidation permet aux établissements de santé de disposer d'une stratégie de gestion des données unique pour tous les workloads Epic et non Epic.

Charges de travail opérationnelles des bases de données

Chaque serveur de base de données Epic effectue des E/S sur les types de fichiers suivants :

- Fichiers de base de données
- Fichiers journaux
- Fichiers d'application

La charge de travail d'un serveur de base de données dépend de son rôle dans l'environnement logiciel Epic. Par exemple, les fichiers de base de données de production sont généralement soumis à la charge de travail la plus exigeante, constituée de 100 % de requêtes en E/S aléatoires. La charge de travail des bases de données en miroir est généralement moins exigeante et présente moins de demandes de lecture. Les

workloads de fichiers journaux sont principalement séquentiels.

Epic maintient un modèle de charge de travail pour le banc d'essai des performances de stockage et la charge de travail des clients. Pour plus d'informations sur le modèle de charge de travail Epic, les résultats d'un banc d'essai et des conseils sur l'utilisation des outils de dimensionnement NetApp pour dimensionner correctement le stockage dans les environnements Epic, voir "Tr-3930i : instructions de dimensionnement NetApp pour Epic" (connexion NetApp requise).

Epic fournit également à chaque client un guide de configuration matérielle personnalisé contenant les projections d'E/S et les besoins en capacité de stockage. Les exigences de stockage finales peuvent inclure des environnements de développement, de test et/ou intermédiaires, ainsi que toute autre charge de travail auxiliaire qui peut être consolidée. Les clients peuvent utiliser le guide de configuration matérielle pour communiquer les besoins totaux en stockage à NetApp. Ce guide contient toutes les données nécessaires au dimensionnement d'un déploiement Epic.

Lors de la phase de déploiement, Epic fournit un guide d'organisation du stockage de base de données, qui fournit des informations plus granulaires au niveau des LUN et peut être utilisé dans le cadre d'une conception de stockage avancée. Notez que le Guide d'organisation du stockage de la base de données est une recommandation générale en matière de stockage et n'est pas spécifique à NetApp. Ce guide vous aidera à déterminer l'infrastructure de stockage la plus adaptée sur NetApp.

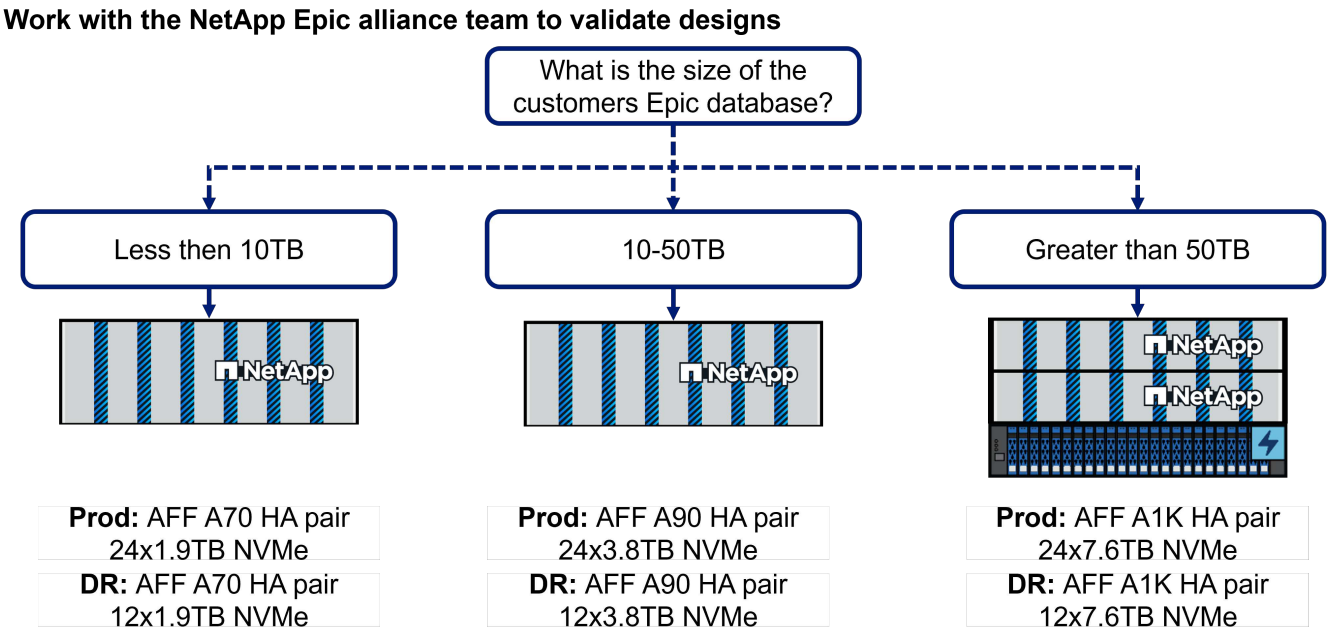
Dimensionnement Epic

Lors du dimensionnement d'un environnement de stockage Epic, la taille de la base de données ODB fait partie des points clés à prendre en compte pour l'architecture.

Vous pouvez utiliser le schéma ci-dessous pour sélectionner une architecture de stockage Epic de petite, moyenne et grande taille. Ces conceptions incluent l'exécution de toutes les charges de travail répertoriées dans le Guide de configuration matérielle. L'arbre de dimensionnement est basé sur les données de plus de 100 guides de configuration matérielle et doit être principalement une estimation précise.

Il est important de noter qu'il ne s'agit que d'un point de départ. Vous devez collaborer avec notre équipe Alliance Epic pour confirmer tous les designs Epic. L'équipe peut être rejointe sur Epic@NetApp.com. Chaque déploiement doit répondre aux demandes des clients tout en respectant les bonnes pratiques recommandées par Epic et NetApp.

- Petite architecture Epic avec une base de données Epic moins de 10 To
- Architecture Epic de taille moyenne avec une base de données Epic de 10 To à 50 To
- Grande architecture Epic avec une base de données Epic de plus de 50 To



Exigences de stockage Epic

Des ressources de stockage dédiées sont généralement fournies pour la base de données de production, tandis que les instances de base de données en miroir partagent des ressources de stockage secondaires avec d'autres composants logiciels Epic, tels que les outils de reporting Clarity.

D'autres environnements de stockage logiciels, tels que ceux utilisés pour les fichiers d'applications et de systèmes, sont également fournis par les ressources de stockage secondaires.

Outre les considérations de dimensionnement, Epic propose les règles d'infrastructure de stockage supplémentaires suivantes et les considérations clés suivantes :

- Depuis 2020, toutes les charges de travail opérationnelles de la base de données (ODB) doivent se trouver sur des baies 100 % Flash.
- Epic recommande que chaque pool de stockage se trouve sur du matériel physique distinct, notamment pool1, pool2, pool3, NAS1 et NAS2.

> Un nœud d'un cluster peut être considéré comme un pool de stockage. Avec ONTAP 9.4 ou version ultérieure et AQoS, vous pouvez créer des pools protégés à l'aide de stratégies.

- Nouvelle recommandation de sauvegarde Epic 3-2-1.
  a. Copie située dans le site distant (reprise après incident)
  b. L'une de ces copies doit se trouver sur une plateforme de stockage différente de la copie principale
  c. Des copies des données

> Les clients qui utilisent NetApp SnapMirror pour la sauvegarde de NetApp ne répondent pas aux recommandations 3-2-1. La raison en est que ONTAP to ONTAP ne satisfait pas à la deuxième exigence indiquée ci-dessus. Vous pouvez utiliser SnapMirror directement depuis ONTAP vers le stockage objet sur site (via StorageGRID, par exemple) ou vers le cloud pour répondre aux exigences d'Epic.

Pour plus d'informations sur les impératifs de stockage, consultez les guides Epic suivants disponibles dans Galaxy :

- Considérations relatives aux SAN
- Produits de stockage et état de la technologie (SPAT)
- Guide de configuration matérielle

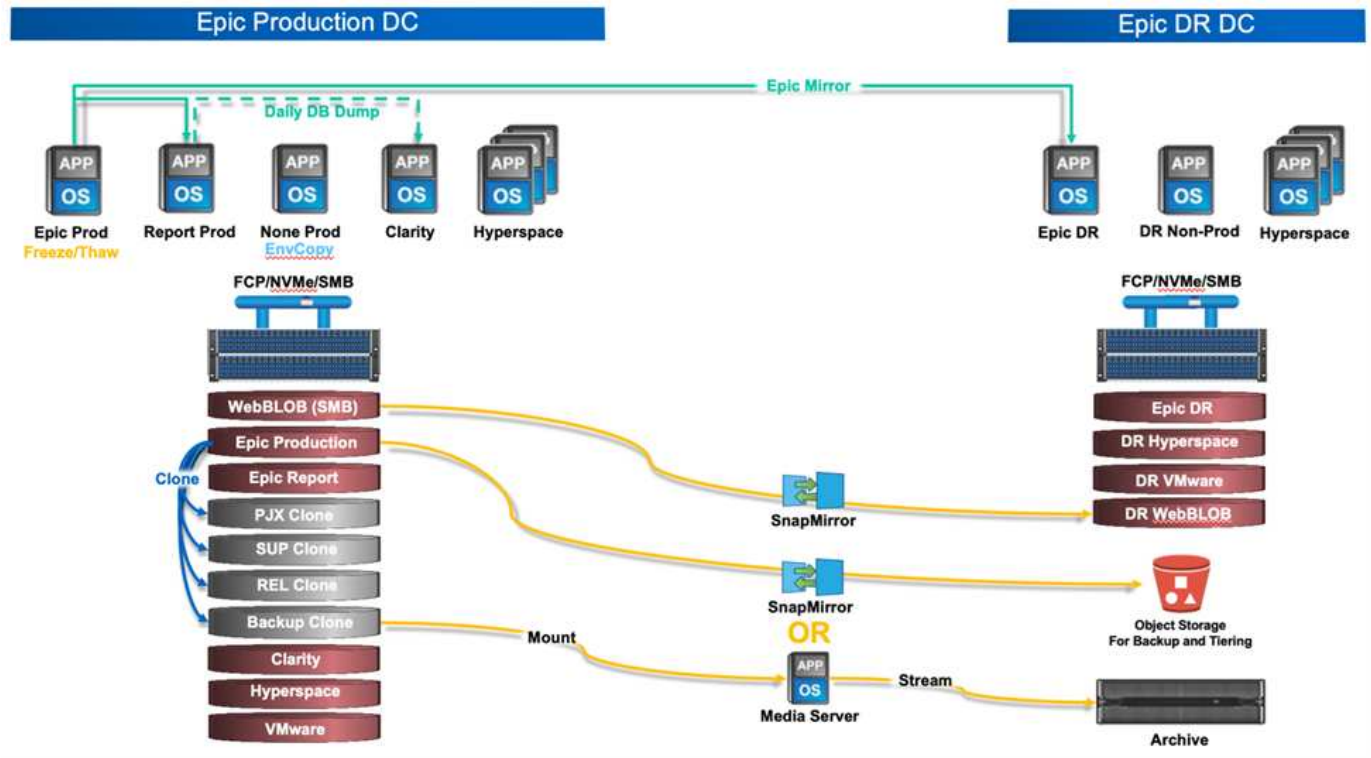
Architecture à quatre nœuds Epic

Les figures ci-dessous illustrent la disposition du stockage pour une architecture à quatre nœuds : une paire haute disponibilité en production et une paire haute disponibilité en reprise après incident. La taille des contrôleurs et le nombre de disques sont basés sur la dernière image de dimensionnement.

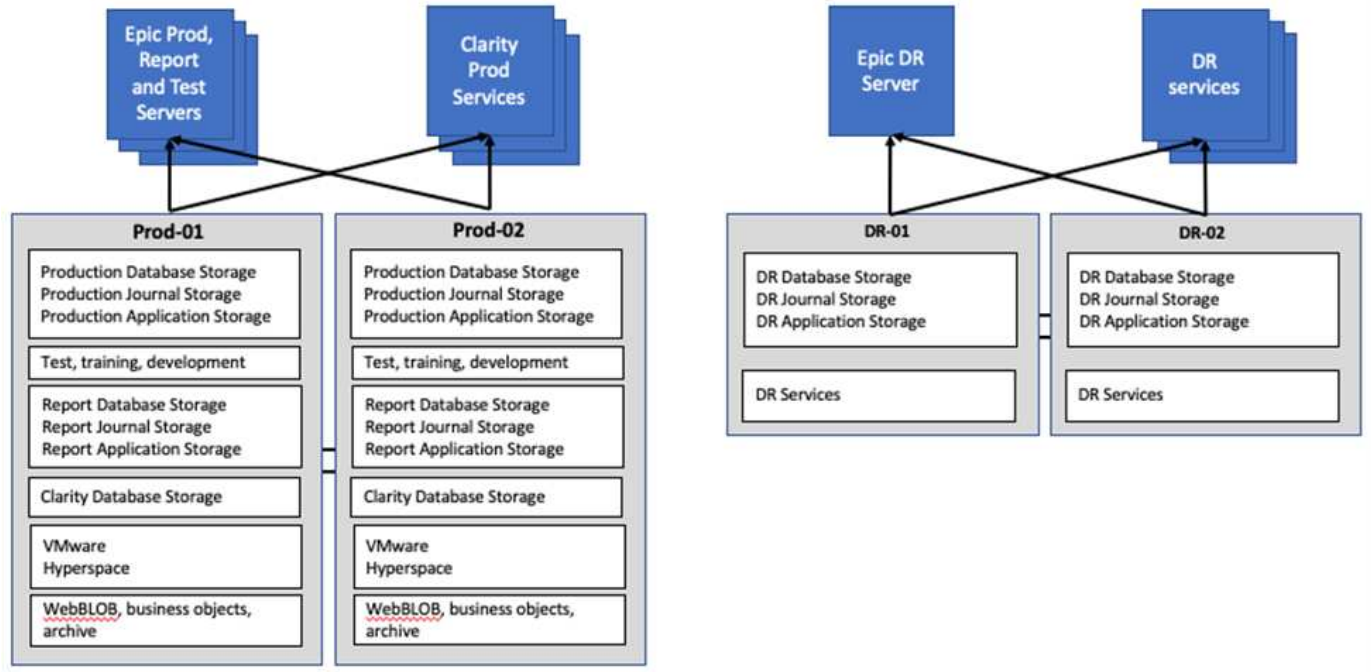
NetApp garantit des performances minimales au niveau du sol en acceptant les politiques de gestion de la qualité de service recommandées par SLM. Epic prend en charge la consolidation des pools de stockage sur ONTAP sur beaucoup moins de matériel. Pour plus d'informations, consultez le document Epic Quarterly SPATS. En fait, pool1, pool2 et NAS1 (répertoriés dans l'Epic Hardware Configuration Guide) peuvent toutes être exécutés sur une seule paire haute disponibilité, les charges de travail étant réparties de manière homogène entre les deux contrôleurs. Dans le cadre de la reprise sur incident, les pools Epic 3 et NAS 3 sont également répartis entre les deux contrôleurs de la paire haute disponibilité.

Les environnements de copie complète de test (SUP, REL et PJX, par exemple) sont clonés à partir d'Epic production, d'Epic Report ou d'Epic Disaster Recovery. Pour plus d'informations sur la sauvegarde et l'actualisation d'Epic, reportez-vous à la section intitulée « gestion des données ».

Architecture à quatre nœuds



Placement des workloads à quatre nœuds

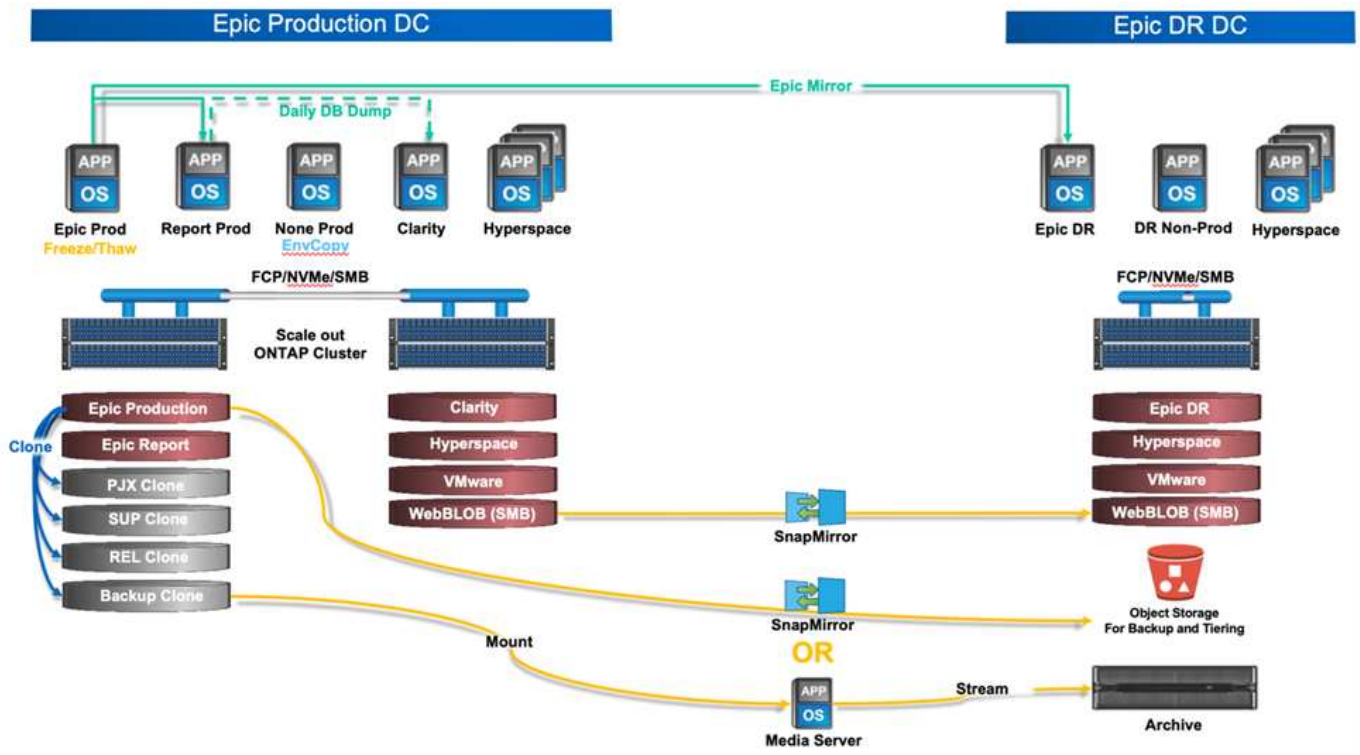


Architecture Epic à 6 nœuds

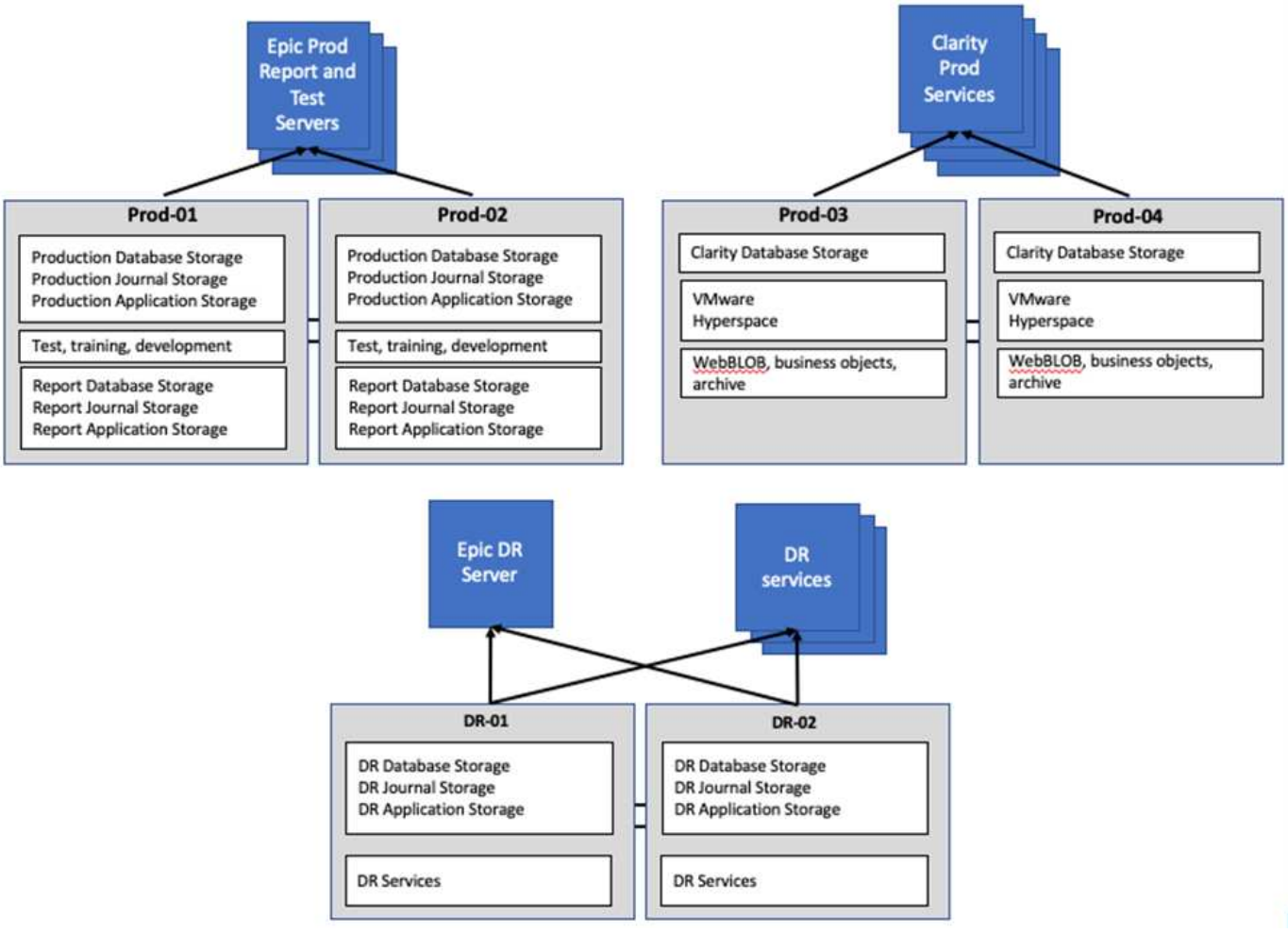
Les clients peuvent commencer par une conception à six nœuds ou évoluer horizontalement de manière fluide de quatre à six nœuds en fonction de la demande croissante. L'évolutivité scale-out permet de déplacer des charges de travail entre les nœuds et de rééquilibrer les opérations sans interruption dans le cluster.

Cette architecture offre le meilleur équilibre entre performances et capacité sur le cluster. Les applications Epic production, Epic Report et Epic Test s'exécutent sur la première paire haute disponibilité. La deuxième paire haute disponibilité est utilisée pour la clarté, l'hyperespace, VMware, NAS1 et les autres workloads Epic. La reprise sur incident est identique à l'architecture à quatre nœuds de la section précédente.

Architecture à six nœuds



Placement des workloads à 6 nœuds

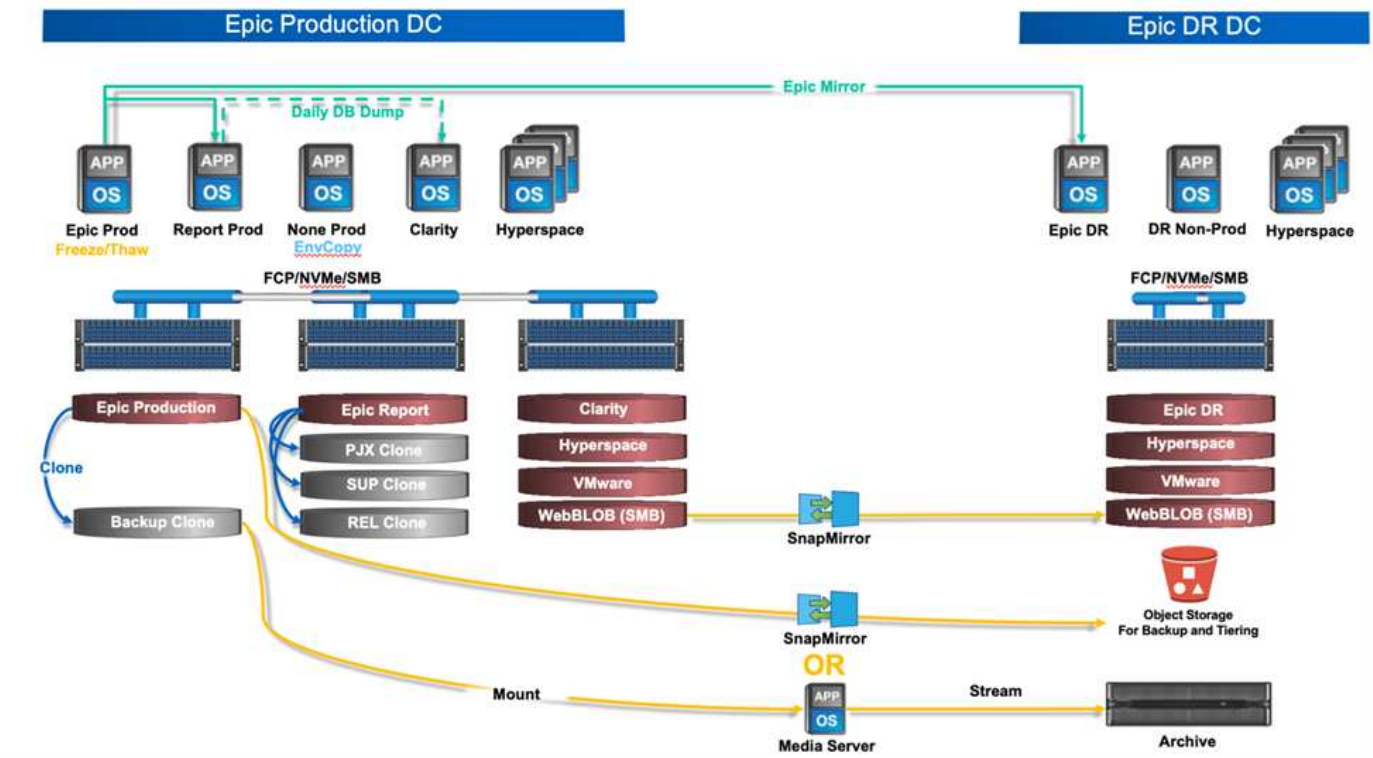


Architecture Epic à huit nœuds

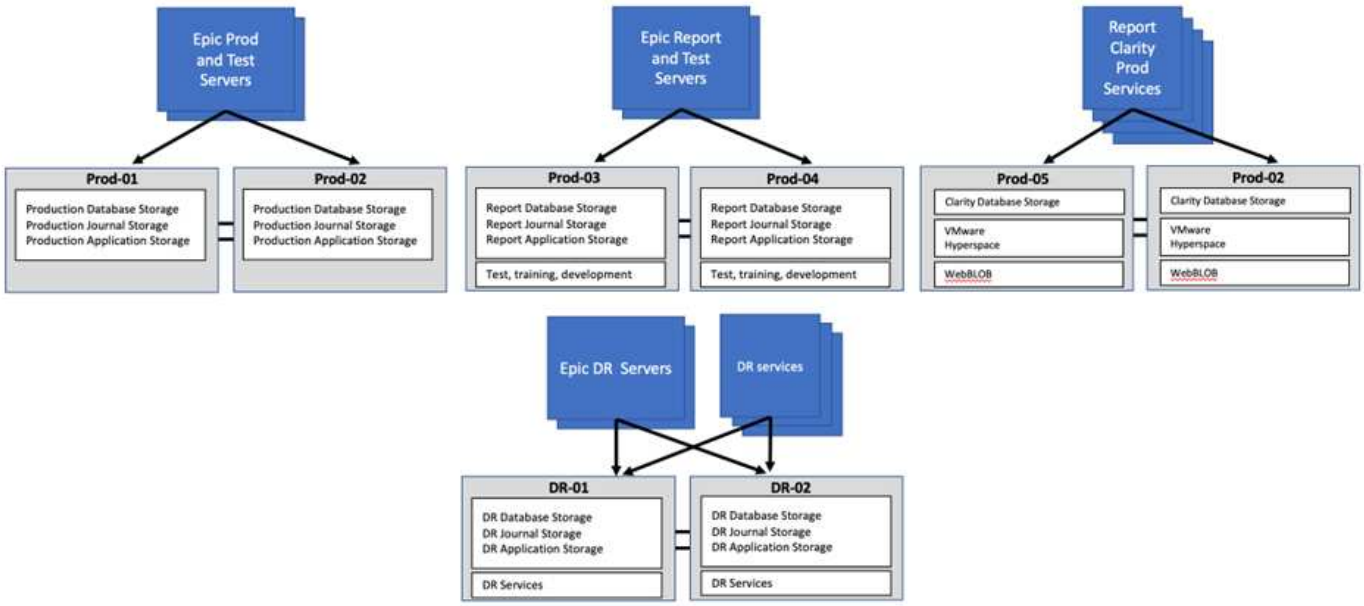
Les figures ci-dessous illustrent l'architecture scale-out à huit nœuds. Là encore, vous pouvez commencer avec quatre nœuds, puis les faire évoluer jusqu'à six, et continuer jusqu'à huit nœuds au-delà. Cette architecture offre le meilleur équilibre entre performance et capacité sur les six nœuds en production.

Dans cette conception, les environnements de test sont clonés à partir du rapport au lieu de la production. La production peut ainsi décharger les environnements de test et les contrôles d'intégrité.

Architecture à huit nœuds



Placement des workloads à huit nœuds



Configuration et meilleures pratiques

Epic sur ONTAP - utilitaires hôtes

Les utilitaires d'hôte NetApp sont des packs logiciels destinés à divers systèmes d'exploitation qui contiennent des utilitaires de gestion tels que `sanlun` le binaire de l'interface de ligne de commande, les pilotes multivoies et d'autres fichiers importants requis pour effectuer correctement des opérations SAN.

NetApp recommande d'installer les utilitaires hôtes NetApp sur les hôtes connectés aux systèmes de stockage NetApp et qui y accèdent. Pour plus d'informations, reportez-vous à la section "Matrice d'interopérabilité" et "Hôtes SAN"à la documentation.

Avec AIX, il est particulièrement important que les utilitaires hôtes soient installés avant de découvrir les LUN. Cela permet de s'assurer que le comportement de chemins d'accès multiples de la LUN est correctement configuré. Si la découverte a été effectuée sans les utilitaires hôtes, les LUN doivent être déconfigurées du système à l'aide de la `rmdev -dl` commande, puis redécouvertes via `cfgmgr` ou un redémarrage.

Configuration de volume et de LUN Epic

Le document recommandations d'Epic Database Storage Layout fournit des conseils sur la taille et le nombre de LUN pour chaque base de données.

Il est important de consulter ce document avec l'aide d'Epic DBA et d'Epic, et de finaliser le nombre de LUN et la taille de LUN en cas d'ajustement. Ces recommandations de stockage sont importantes pour la profondeur de file d'attente des HBA, les performances de stockage, la facilité d'exploitation et la facilité d'extension.

Pour connaître la profondeur de la file d'attente du système d'exploitation du serveur, utilisez au moins huit LUN (une LUN par volume) pour une base de données. Augmentez le nombre de LUN en fonction du nombre de nœuds dans le cluster ONTAP. Par exemple, ajoutez 4 LUN si vous utilisez un cluster à 4 nœuds (2 paires HA). Pour les environnements de plus grande taille, davantage de LUN peuvent être nécessaires ; utilisez le même nombre de volumes (huit au total, distribués sur un nœud de stockage) et ajoutez des LUN par multiples de deux entre les nœuds et les volumes du cluster. Cette approche vous permet de faire évoluer facilement votre environnement Epic.

Exemple 1 : cluster ONTAP à 2 nœuds

2 nœuds, 1 paire HA 8 volumes, 4 volumes par nœud 8 LUN, une LUN par volume Ajout de 2 LUN supplémentaires, une LUN sur le nœud 01 dans le volume 01, une LUN sur le nœud 02 dans le volume 02.

Exemple 2 : cluster ONTAP à 4 nœuds

4 nœuds, 2 paires HA 8 volumes, 2 volumes par nœud 8 LUN, une LUN par volume Ajout de 4 LUN supplémentaires, une LUN sur le nœud 01 dans le volume01, une LUN sur le nœud 02 dans le volume02, une LUN sur le nœud 03 dans le volume03, une LUN sur le nœud 04 dans le volume04.

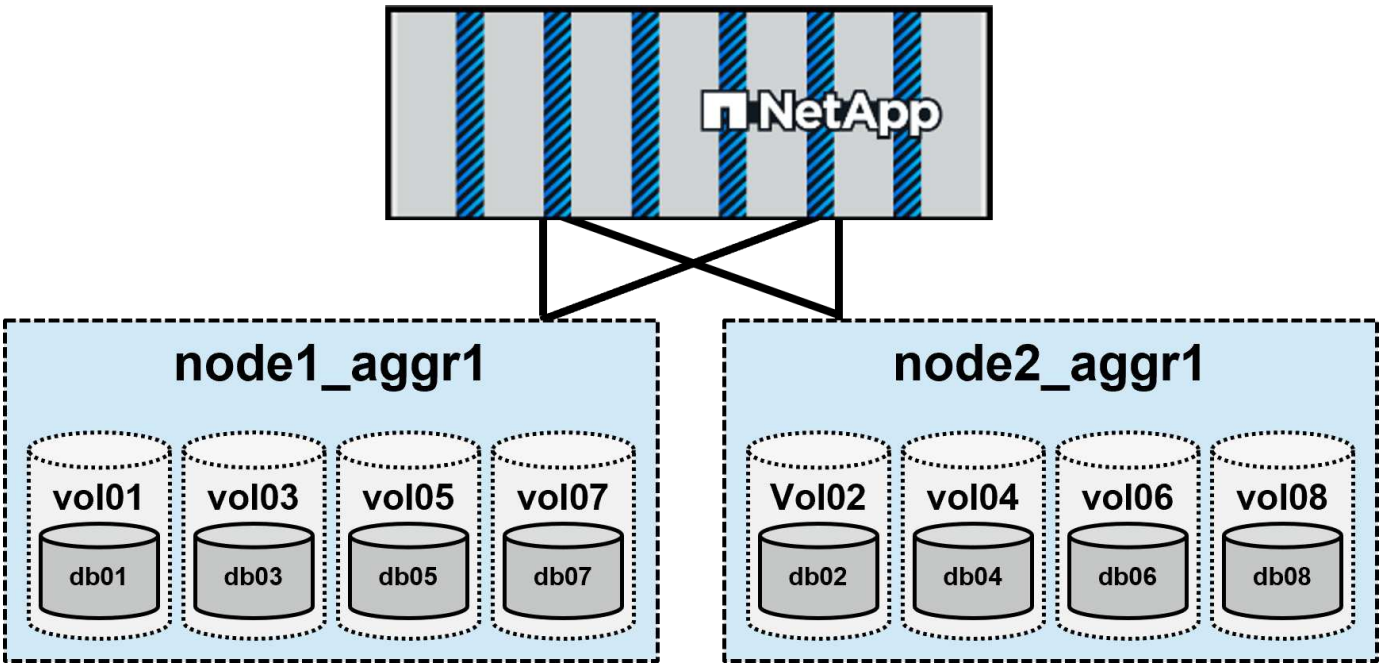
Pour optimiser les performances d'une charge de travail, comme Epic ODB ou Clarity, chaque disposition est également optimisée pour le stockage NetApp. Avec huit volumes utilisés, les E/S d'écriture sont réparties de façon homogène entre les contrôleurs, optimisant ainsi l'utilisation du processeur. Pour la réplication et la sauvegarde, il est préférable de limiter le nombre de volumes à huit pour simplifier les opérations.

Options d'évolutivité

Si le serveur requiert plus de stockage, l'option la plus simple est d'augmenter les LUN contenant des volumes. La seconde option consiste à ajouter des LUN aux groupes de volumes par multiples de deux à la fois (un par volume par nœud).

Exemple :

Disposition des volumes et 8 LUN



Si dans un grand environnement nécessitant plus de 4 nœuds ou 8 LUN, consultez notre équipe Epic pour confirmer les conceptions de LUN. L'équipe peut être rejointe sur Epic@NetApp.com.

Et des meilleures pratiques

- Utilisez 8 LUN dans 8 volumes pour démarrer, en ajoutant 2 LUN à la fois, sur tous les nœuds du cluster.
- Équilibrez les charges de travail sur toute la paire haute disponibilité afin d'optimiser les performances et l'efficacité.
- Créez des LUN dont la taille est prévue pour 3 ans de croissance. (Pour connaître la taille maximale de LUN, consultez la"Documentation de l'ONTAP".)
- Utilisation de volumes et de LUN à provisionnement fin
- Utilisez au moins huit LUN de base de données, deux LUN de journal et deux LUN d'application. Cette configuration optimise les performances du stockage et la profondeur de la file d'attente du système d'exploitation. Vous pouvez utiliser davantage de ressources si nécessaire pour des raisons de capacité ou autres.
- Si vous devez ajouter des LUN à des groupes de volumes, ajoutez huit LUN à la fois.
- Les groupes de cohérence sont requis pour la sauvegarde conjointe du groupe de volumes et des LUN.
- N'utilisez pas QoS pendant le Genio ou les performances d'E/S.
- Après les tests Genio ou Clarity, NetApp recommande de supprimer le stockage et de le reprovisionner avant de charger les données de production.
- Il est important que `-space-allocation` l'option activé soit définie sur les LUN. Si ce n'est pas le cas, les données supprimées des LUN ne sont pas visibles par ONTAP et peuvent entraîner des problèmes de capacité. Pour plus d'informations, consultez le guide de référence rapide de la configuration du stockage Epic.

Protocoles Epic et de fichiers

Il est possible d'associer NAS et SAN au sein d'une même baie 100 % Flash.

> **NetApp recommande** d'utiliser des volumes FlexGroup pour les partages NAS, tels que WebBLOB (si disponible).

WebBLOB représente jusqu'à 95 % des données inactives. Vous pouvez également libérer de l'espace sur votre baie 100 % Flash et procéder au Tiering des sauvegardes et des données inactives vers un stockage objet sur site ou dans le cloud à l'aide de la "FabricPool"fonctionnalité ONTAP. Tout cela peut être réalisé sans effet notable sur les performances. FabricPool est une fonctionnalité incluse dans ONTAP. Les clients peuvent générer un rapport de données inactives pour évaluer les bénéfices potentiels grâce à l'activation de FabricPool. Vous pouvez définir l'âge des données à hiérarchiser via une règle. Les clients Epic ont réalisé des économies considérables grâce à cette fonctionnalité.

Gestion des performances Epic

La plupart des baies 100 % Flash offrent les performances requises pour les workloads Epic. L'atout de NetApp est sa capacité à définir des règles de performance au niveau du sol et à garantir une performance prévisible pour chaque application.

Qualité de service (QoS)

NetApp recommande d'utiliser la QoS. Le bénéfice de la QoS est la possibilité de consolider tous les workloads Epic. Tous les protocoles et les pools de stockage peuvent résider sur moins de matériel. Il n'est pas nécessaire de séparer les pools de stockage.

- NetApp recommande d'attribuer à toutes les charges de travail du cluster une politique de qualité de service afin de mieux gérer la marge sur le cluster.
- NetApp recommande d'équilibrer toutes les charges de travail de façon homogène dans la paire haute disponibilité.
- N'utilisez pas les règles de qualité de service lors des tests d'E/S ; dans le cas contraire, les tests Genio échoueront. Analysez les différents workloads de production pendant 2-4 semaines avant d'attribuer des règles de QoS.

Epic sur ONTAP - protocoles

FCP est le protocole privilégié pour la présentation des LUN.

> **NetApp recommande** une segmentation à un seul initiateur : un initiateur par zone avec tous les ports cibles requis sur le stockage à l'aide des WWPN (Worldwide Port Name). La présence de plusieurs initiateurs dans une même zone est susceptible d'entraîner une diaphonie intermittente des HBA, ce qui provoque des perturbations importantes.

Une fois la LUN créée, mappez-la sur le groupe initiateur (igroup) contenant les WWPN de l'hôte pour permettre l'accès.

NetApp prend également en charge l'utilisation de NVMe/FC (si certaines versions des systèmes d'exploitation AIX et RHEL sont compatibles) et améliore les performances. FCP et NVMe/FC peuvent coexister sur la même structure.

Configuration de l’efficacité du stockage Epic

Les fonctionnalités d’efficacité à la volée de ONTAP sont activées par défaut et fonctionnent indépendamment du protocole, de l’application ou du Tier de stockage.

Les fonctionnalités d’efficacité réduisent la quantité de données écrites sur un système de stockage Flash coûteux et le nombre de disques requis. ONTAP préserve l’efficacité de la réplication. L’efficacité n’a que peu, voire pas du tout d’effet sur les performances, même pour une application sensible à la latence comme Epic.

NetApp recommande d’activer tous les paramètres d’efficacité pour optimiser l’utilisation du disque. Ces paramètres sont activés par défaut sur les systèmes AFF et ASA.

Les fonctionnalités suivantes rendent ce stockage efficace possible :

- La déduplication économise de l’espace sur le stockage primaire en supprimant les copies redondantes des blocs d’un volume qui héberge des LUN. Cette option recommandée est activée par défaut.
- La compression à la volée réduit la quantité de données à écrire sur le disque, et les workloads Epic réalisent des économies d’espace considérables. Cette option recommandée est activée par défaut.
- La compaction à la volée permet de combiner des blocs de 4 ko moins de moitié pleins dans un seul bloc. Cette option recommandée est activée par défaut.
- La réplication fine est au cœur de la gamme de logiciels de protection des données NetApp, qui inclut le logiciel NetApp SnapMirror. La réplication fine de SnapMirror protège les données stratégiques tout en limitant les besoins en capacité de stockage. **NetApp recommande** d’activer cette option.
- Déduplication dans l’agrégat. La déduplication a toujours été au niveau du volume. Avec ONTAP 9.2, la déduplication de l’agrégat est désormais disponible, ce qui permet de réaliser des économies supplémentaires en termes de réduction de disque. La déduplication post-traitement dans l’agrégat a été ajoutée à ONTAP 9.3. **NetApp recommande** d’activer cette option.

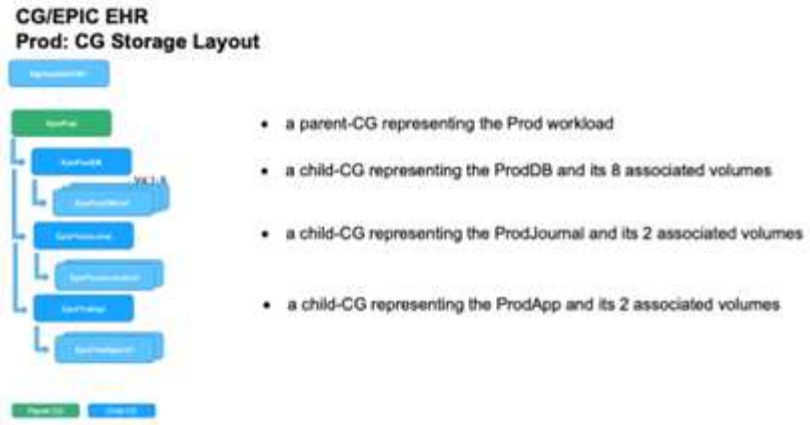
Configuration de l’efficacité du stockage Epic

Les applications avec stockage réparti sur plusieurs volumes avec une ou plusieurs LUN de quantité appropriée à la charge de travail nécessitent que le contenu soit sauvegardé ensemble pour assurer la cohérence de la protection des données.

Les groupes de cohérence (CGS pour Short) proposent cette fonctionnalité et bien plus encore. Elles peuvent être utilisées chaque nuit pour créer des copies Snapshot cohérentes à la demande ou planifiées à l’aide d’une règle. Vous pouvez l’utiliser pour restaurer, cloner et même répliquer des données.

Pour plus d’informations sur CGS, reportez-vous au "Présentation des groupes de cohérence"

Une fois les volumes et les LUN provisionnés comme détaillé dans les sections précédentes de ce document, ils peuvent ensuite être configurés dans un ensemble de groupes de cohérence. La meilleure pratique recommandée est de les configurer comme indiqué dans l’image ci-dessous :



Snapshots de groupes de cohérence

Un planning de snapshots de groupe de cohérence doit être défini chaque jour sur les groupes de cohérence enfant associés aux volumes fournissant du stockage pour la base de données de production. Cela donnera lieu à un nouvel ensemble de sauvegardes cohérentes de ces CGS chaque nuit. Vous pouvez ensuite les utiliser pour cloner la base de données de production afin de les utiliser dans des environnements hors production, tels que le développement et le test. NetApp a développé des workflows Ansible automatisés basés sur un groupe de cohérence propriétaire pour Epic afin d'automatiser la sauvegarde des bases de données de production, l'actualisation et les environnements de test.

Les snapshots de groupe de cohérence peuvent être utilisés pour prendre en charge les opérations de restauration de la base de données de production d'Epic.

Pour les volumes SAN, désactivez la règle de snapshot par défaut sur chaque volume utilisé pour les groupes de cohérence. Ces snapshots sont généralement gérés par l'application de sauvegarde utilisée ou le service d'automatisation Epic Ansible de NetApp.

Pour les volumes SAN, désactivez la règle de snapshot par défaut sur chaque volume. Ces snapshots sont généralement gérés par une application de sauvegarde ou par l'automatisation Ansible d'Epic.[NS2]

Les datasets WebBLOB et VMware doivent être configurés comme des volumes uniquement, et non associés à CGS. Vous pouvez utiliser SnapMirror pour conserver les snapshots sur des systèmes de stockage distincts de la production.

Une fois la configuration terminée, elle se présente comme suit :

Dimensionnement du stockage pour Epic

Vous devez collaborer avec notre équipe Alliance Epic pour confirmer tous les designs Epic. L'équipe peut être rejointe sur Epic@NetApp.com. Chaque déploiement doit répondre aux demandes des clients tout en respectant les bonnes pratiques recommandées par Epic et NetApp.

Pour plus d'informations sur l'utilisation des outils de dimensionnement NetApp afin de déterminer la taille et le nombre corrects de groupes RAID pour les besoins de stockage de l'environnement logiciel Epic, reportez-vous à la section "Tr-3930i : instructions de dimensionnement NetApp pour Epic" (connexion NetApp requise).

Vous devez disposer d'un accès au Field Portal NetApp.

Informations supplémentaires sur Epic sur ONTAP

Pour en savoir plus sur les informations fournies dans ce document, consultez ces documents et/ou sites web :

- "Documentation produit NetApp"
- "Documentation ONTAP 9"
- "Groupes de cohérence"
- "Ressources de documentation du gestionnaire de système ONTAP et ONTAP"
- "Tr-3930i : instructions de dimensionnement NetApp pour Epic" (Connexion NetApp requise)

Documents d'orientation client Epic

Epic fournit aux clients les documents suivants à des fins d'assistance sur les serveurs, le stockage et le réseau. Ces documents sont référencés dans ce rapport technique.

- Considérations relatives au réseau de stockage
- Guide des solutions techniques de continuité de l'activité
- Guide de référence sur la stratégie d'architecture 100 % Flash
- Produits de stockage et état de la technologie
- Considérations relatives au cloud EPIC
- Guide de configuration matérielle (spécifique au client)
- Recommandations relatives à l'infrastructure de stockage de la base de données (spécifique au client)

Hyper-V

Instructions de déploiement et meilleures pratiques de stockage

Présentation

Microsoft Windows Server est un système d'exploitation (OS) professionnel qui couvre la mise en réseau, la sécurité, la virtualisation, le cloud privé, le cloud hybride, infrastructure de postes de travail virtuels, protection des accès, protection des informations, services web, infrastructure de plate-forme applicative, et bien plus encore.

> **Cette documentation remplace les rapports techniques publiés précédemment *TR-4568: Consignes de déploiement NetApp et meilleures pratiques de stockage pour Windows Server***

ONTAP s'exécute sur les contrôleurs de stockage NetApp. Il est disponible dans plusieurs formats.

- Architecture unifiée prenant en charge les protocoles de fichiers, d'objets et de blocs. Les contrôleurs de stockage peuvent ainsi agir en tant que périphériques NAS et SAN, ainsi qu'en tant que magasins d'objets
- Une baie 100 % SAN (ASA) axée uniquement sur les protocoles de niveau bloc et qui optimise les temps de reprise des E/S (IORT) en ajoutant un chemins d'accès multiples actif-actif symétrique pour les hôtes connectés
- Architecture unifiée Software-defined
  - ONTAP Select s'exécutant sur VMware vSphere ou KVM
  - Cloud Volumes ONTAP s'exécutant en tant qu'instance cloud native
- Offres propriétaires de fournisseurs de cloud hyper-évolutif
  - Amazon FSX pour NetApp ONTAP
  - Azure NetApp Files
  - Google Cloud NetApp volumes

ONTAP offre des fonctionnalités d'efficacité du stockage NetApp, telles que la technologie Snapshot de NetApp, le clonage, la déduplication, le provisionnement fin, la réplication fine, la compression, la hiérarchisation du stockage virtuel, et bien plus encore, avec des performances et une efficacité améliorées.

Ensemble, Windows Server et ONTAP peuvent fonctionner dans de grands environnements et apporter une valeur considérable à la consolidation des data centers et aux déploiements de clouds privés ou hybrides. Cette combinaison assure également des charges de travail efficaces sans interruption et assure une évolutivité transparente.

Public visé

Ce document est destiné aux architectes système et de stockage qui conçoivent des solutions de stockage NetApp pour Windows Server.

Nous faisons les hypothèses suivantes dans ce document :

- Le lecteur a une connaissance générale des solutions matérielles et logicielles NetApp. Voir la "Guide

d'administration du système destiné aux administrateurs du cluster" pour plus d'informations.

- Le lecteur possède des connaissances générales sur les protocoles d'accès aux blocs, tels que iSCSI, FC et le protocole d'accès aux fichiers SMB/CIFS. Voir la "Gestion de l'environnement SAN clustered Data ONTAP" Pour obtenir des informations sur SAN. Voir la "Gestion NAS" Pour des informations relatives à CIFS/SMB.
- Le lecteur a des connaissances générales sur le système d'exploitation Windows Server et Hyper-V.

Pour obtenir une matrice complète et régulièrement mise à jour des configurations SAN et NAS testées et prises en charge, consultez le sur le "Matrice d'interopérabilité (IMT)" site de support NetApp. Avec IMT, vous pouvez déterminer les versions de produits et de fonctionnalités prises en charge pour votre environnement spécifique. Le NetApp IMT définit les composants et versions du produit compatibles avec les configurations prises en charge par NetApp. Les résultats dépendent des installations de chaque client et de leur conformité aux spécifications publiées.

Stockage NetApp et environnement Windows Server

Comme indiqué dans le "Présentation", Les contrôleurs de stockage NetApp offrent une architecture réellement unifiée qui prend en charge les protocoles de fichiers, de blocs et d'objets. Notamment SMB/CIFS, NFS, NVMe/TCP, NVMe/FC, iSCSI, FC(FCP) et S3 créent un accès client et hôte unifié. Le même contrôleur de stockage peut fournir simultanément un service de stockage bloc sous la forme de LUN SAN et de service de fichiers comme NFS et SMB/CIFS. ONTAP est également disponible en tant que baie 100 % SAN (ASA) qui optimise l'accès à l'hôte via des chemins d'accès multiples symétriques actif-actif avec iSCSI et FCP, tandis que les systèmes ONTAP unifiés utilisent des chemins d'accès multiples asymétriques actifs/actifs. Dans les deux modes, ONTAP utilise ANA pour la gestion des chemins d'accès multiples NVMe over Fabrics (NVMe-of).

Un contrôleur de stockage NetApp exécutant ONTAP peut prendre en charge les charges de travail suivantes dans un environnement Windows Server :

- Machines virtuelles hébergées dans des partages SMB 3.0 disponibles en continu
- Serveurs virtuels hébergés sur des LUN CSV (Cluster Shared Volume) s'exécutant sur iSCSI ou FC
- Bases de données SQL Server sur partages SMB 3.0
- Bases de données SQL Server sur NVMe-of, iSCSI ou FC
- Autres charges de travail applicatives

De plus, les fonctionnalités d'efficacité du stockage NetApp, telles que la déduplication, les copies NetApp FlexClone®, la technologie Snapshot NetApp, le provisionnement fin, la compression et la hiérarchisation du stockage offrent des avantages significatifs pour les charges de travail exécutées sur Windows Server.

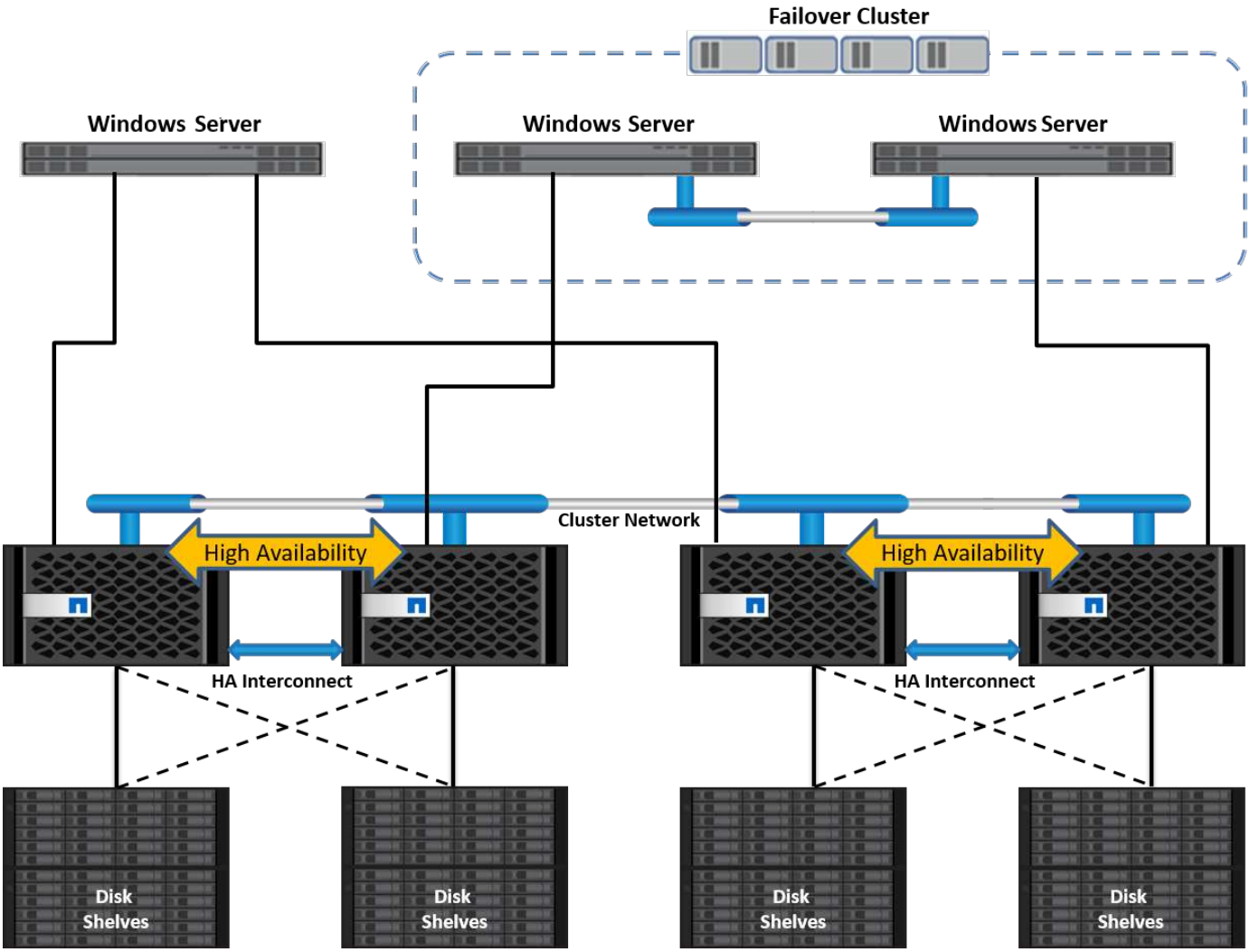
Gestion des données ONTAP

ONTAP est un logiciel de gestion qui s'exécute sur un contrôleur de stockage NetApp. Appelé nœud, le contrôleur de stockage NetApp est un périphérique matériel doté d'un processeur, d'une mémoire RAM et d'une mémoire NVRAM. Le nœud peut être connecté à des disques SATA, SAS ou SSD, ou à une combinaison de ces disques.

Plusieurs nœuds sont agrégés dans un système en cluster. Les nœuds du cluster communiquent entre eux de

manière continue pour coordonner les activités du cluster. Les nœuds peuvent également déplacer les données de manière transparente d'un nœud à l'autre à l'aide de chemins redondants vers une interconnexion de cluster dédiée constituée de deux commutateurs Ethernet 10 Gbit/s. Les nœuds du cluster peuvent se prendre en charge pour assurer une haute disponibilité dans tous les scénarios de basculement. Les clusters sont administrés dans l'ensemble du cluster plutôt que par nœud, tandis que les données sont servies depuis un ou plusieurs serveurs virtuels de stockage (SVM). Un cluster doit disposer d'au moins un SVM pour assurer le service des données.

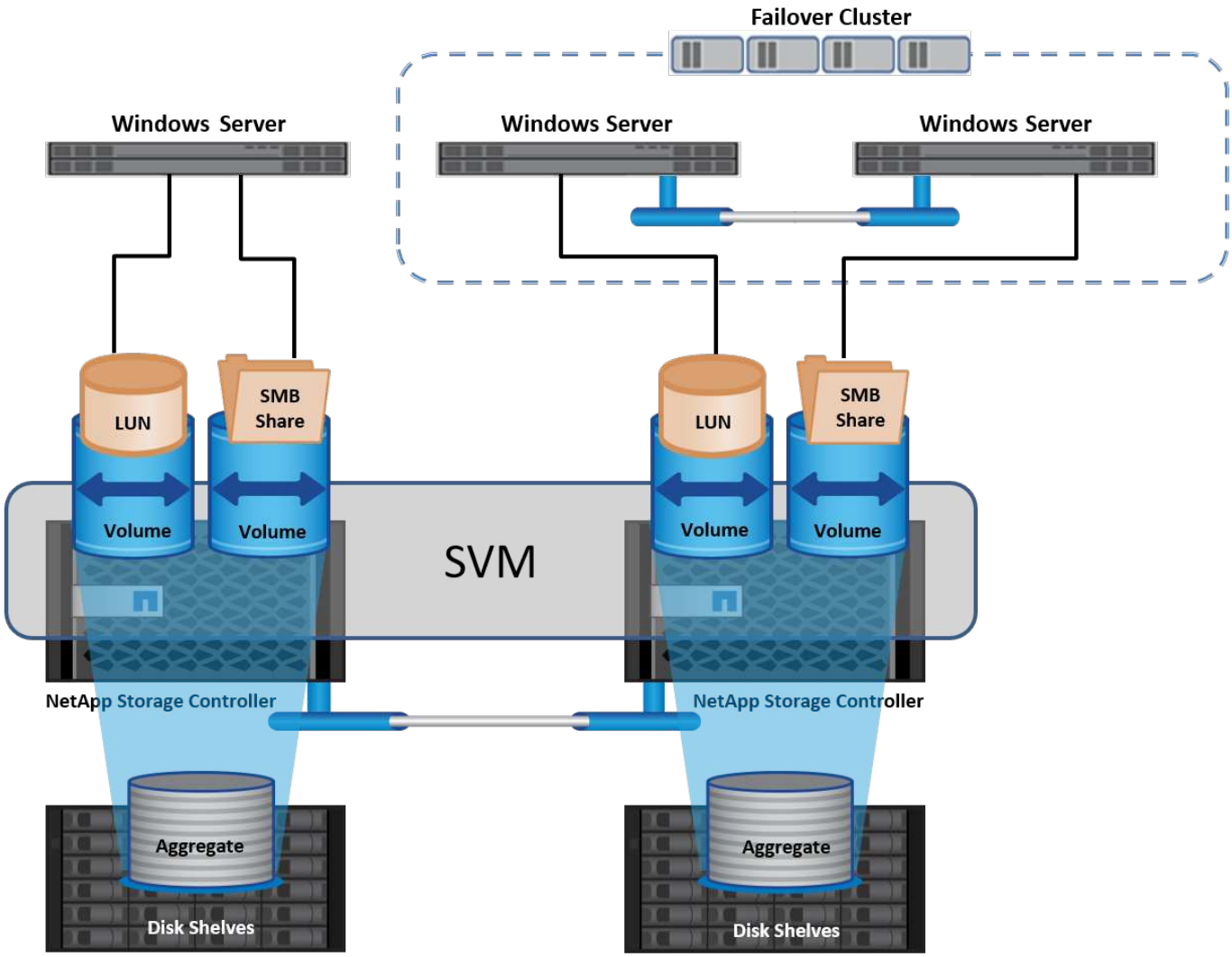
L'unité de base d'un cluster est le nœud, et des nœuds sont ajoutés au cluster dans le cadre d'une paire haute disponibilité. Les paires HAUTE DISPONIBILITÉ assurent une haute disponibilité en communiquant les unes avec les autres via une interconnexion haute disponibilité (distincte de l'interconnexion de cluster dédiée) et en maintenant des connexions redondantes aux disques de la paire haute disponibilité. Les disques ne sont pas partagés entre les paires haute disponibilité, bien que les tiroirs puissent contenir des disques appartenant à l'un ou l'autre des membres d'une paire haute disponibilité. La figure suivante illustre le déploiement d'un système de stockage NetApp dans un environnement Windows Server.



Ordinateurs virtuels de stockage

Un SVM ONTAP est un serveur de stockage logique qui permet d'accéder aux données des LUN et/ou d'un espace de noms NAS à partir d'une ou plusieurs interfaces logiques (LIF). La SVM est donc l'unité de base de segmentation du stockage qui permet une colocation sécurisée dans ONTAP. Chaque SVM est configuré pour ses propres volumes de stockage provisionnés à partir d'un agrégat physique et d'interfaces logiques (LIF) attribuées à un réseau Ethernet physique ou à des ports cibles FC.

Des disques logiques (LUN) ou des partages CIFS sont créés au sein des volumes d'un SVM et sont mappés sur des hôtes et des clusters Windows afin de leur fournir de l'espace de stockage, comme illustré dans la figure ci-dessous. Les SVM sont indépendants des nœuds et basés sur le cluster. Ils peuvent utiliser des ressources physiques telles que des volumes ou des ports réseau n'importe où dans le cluster.



Provisionnement du stockage NetApp pour Windows Server

Le stockage peut être provisionné sur Windows Server dans les environnements SAN et NAS. Dans un environnement SAN, le stockage est fourni sous forme de disques provenant de LUN sur un volume NetApp en tant que stockage en mode bloc. Dans un environnement NAS, le stockage est fourni sous forme de partages CIFS/SMB sur des volumes NetApp comme stockage de fichiers. Ces disques et partages peuvent être appliqués dans Windows Server comme suit :

- Stockage pour les hôtes Windows Server pour les charges de travail applicatives
- Stockage pour Nano Server et conteneurs
- Stockage des hôtes Hyper-V individuels pour stocker les machines virtuelles
- Stockage partagé pour les clusters Hyper-V sous la forme de CSV pour le stockage des machines virtuelles
- Stockage pour bases de données SQL Server

Gestion du stockage NetApp

Pour connecter, configurer et gérer le stockage NetApp à partir de Windows Server 2016, utilisez l'une des méthodes suivantes :

- **Secure Shell (SSH).** utilisez n'importe quel client SSH sur Windows Server pour exécuter les commandes CLI de NetApp.
- **System Manager.** il s'agit du produit de gestion basé sur l'interface graphique de NetApp.
- **Boîte à outils PowerShell NetApp.** il s'agit du kit d'outils PowerShell NetApp pour l'automatisation et la mise en œuvre de scripts et de flux de travail personnalisés.

Kit NetApp PowerShell

Le kit NetApp PowerShell (PSTK) est un module PowerShell qui offre une automatisation de bout en bout et permet d'administrer le stockage ONTAP. Le module ONTAP contient plus de 2,000 cmdlets et aide à l'administration de NetApp AFF, FAS, du matériel générique et des ressources cloud.

Choses à retenir

- NetApp ne prend pas en charge les espaces de stockage Windows Server. Les espaces de stockage sont utilisés uniquement pour JBOD (une simple concaténation de disques) et ne fonctionnent avec aucun type RAID (stockage DAS ou SAN).
- Les pools de stockage en cluster dans Windows Server ne sont pas pris en charge par ONTAP.
- NetApp prend en charge le format de disque dur virtuel partagé (VHDX) pour une mise en cluster invitée dans les environnements SAN Windows.
- Windows Server ne prend pas en charge la création de pools de stockage à l'aide de LUN iSCSI ou FC.

Lecture ultérieure

- Pour plus d'informations sur le kit NetApp PowerShell, rendez-vous sur le "Site de support NetApp".
- Pour plus d'informations sur les bonnes pratiques du kit NetApp PowerShell, reportez-vous à la section "Tr-4475 : guide des bonnes pratiques du kit NetApp PowerShell".

Meilleures pratiques en matière de mise en réseau

Les réseaux Ethernet peuvent être répartis de manière large en plusieurs groupes :

- Un réseau client pour les VM
- Un autre réseau de stockage (connexion iSCSI ou SMB aux systèmes de stockage)
- Un réseau de communication en cluster (battement de cœur et autre communication entre les nœuds du cluster)
- Un réseau de gestion (pour surveiller et dépanner le système)
- Un réseau de migration (pour la migration en direct des hôtes)
- Réplication de machine virtuelle (réplication Hyper-V)

Et des meilleures pratiques

- NetApp recommande de disposer de ports physiques dédiés à chacune des fonctionnalités précédentes pour l'isolation du réseau et les performances.

- Pour chacune des exigences réseau précédentes (à l'exception des exigences de stockage), plusieurs ports réseau physiques peuvent être agrégés pour répartir la charge ou fournir une tolérance aux pannes.
- NetApp recommande de créer un commutateur virtuel dédié sur l'hôte Hyper-V pour la connexion au stockage invité au sein de la machine virtuelle.
- Assurez-vous que les chemins de données iSCSI de l'hôte Hyper-V et de l'invité utilisent différents ports physiques et commutateurs virtuels pour une isolation sécurisée entre l'invité et l'hôte.
- NetApp recommande d'éviter le regroupement de cartes réseau pour les cartes réseau iSCSI.
- NetApp recommande d'utiliser le protocole MPIO (ONTAP Multipath Input/Output) configuré sur l'hôte à des fins de stockage.
- NetApp recommande d'utiliser MPIO sur une machine virtuelle invitée si des initiateurs iSCSI invités sont utilisés. L'utilisation de MPIO doit être évitée au sein de l'invité si vous utilisez des disques directs. Dans ce cas, l'installation de MPIO sur l'hôte devrait suffire.
- NetApp recommande de ne pas appliquer de règles de qualité de service au commutateur virtuel attribué au réseau de stockage.
- NetApp recommande de ne pas utiliser l'adressage IP privé automatique (APIPA) sur les cartes réseau physiques car APIPA n'est pas routable et n'est pas enregistré dans le DNS.
- NetApp recommande d'activer les trames Jumbo pour les réseaux CSV, iSCSI et de migration dynamique afin d'augmenter le débit et de réduire les cycles du processeur.
- NetApp recommande de décocher l'option Autoriser le système d'exploitation de gestion à partager cette carte réseau pour que le commutateur virtuel Hyper-V crée un réseau dédié pour les machines virtuelles.
- NetApp recommande de créer des chemins réseau redondants (plusieurs commutateurs) pour la migration en direct et le réseau iSCSI pour assurer la résilience et la qualité de service.

Le provisionnement dans des environnements SAN

Les SVM ONTAP prennent en charge les protocoles de niveau bloc iSCSI et FC. Lorsqu'un SVM est créé avec un protocole de bloc iSCSI ou FC, le SVM obtient respectivement un nom qualifié iSCSI (IQN) ou un nom WWN FC. Cet identifiant présente une cible SCSI aux hôtes qui accèdent au stockage en bloc NetApp.

Provisionnement de LUN NetApp sur Windows Server

Prérequis

L'utilisation d'un stockage NetApp dans des environnements SAN sous Windows Server présente les conditions suivantes :

- Un cluster NetApp est configuré avec un ou plusieurs contrôleurs de stockage NetApp.
- Le cluster NetApp ou les contrôleurs de stockage disposent d'une licence iSCSI valide.
- Des ports configurés iSCSI et/ou FC sont disponibles.
- La segmentation FC est effectuée sur un commutateur FC pour FC.
- Au moins un agrégat est créé.
- Un SVM doit avoir une LIF par réseau Ethernet ou une structure Fibre Channel sur chaque contrôleur de stockage qui va transmettre des données via iSCSI ou Fibre Channel.

Déploiement

1. Créez un SVM avec le protocole de bloc iSCSI et/ou FC activé. Il est possible de créer un SVM avec l'une des méthodes suivantes :
   - Commandes CLI sur le stockage NetApp
   - ONTAP System Manager
   - Kit NetApp PowerShell
2. Configuration du protocole iSCSI et/ou FC
3. Assigner le SVM avec des LIFs sur chaque nœud de cluster.
4. Démarrer le service iSCSI et/ou FC sur le SVM

   .

5. Créez des datasets de ports iSCSI et/ou FC à l'aide des LIF du SVM.
6. Créez un groupe initiateur iSCSI et/ou FC pour Windows à l'aide du jeu de ports créé.
7. Ajouter un initiateur au groupe initiateur. L'initiateur est l'IQN pour iSCSI et WWPN pour FC. Ils peuvent être interrogés à partir de Windows Server en exécutant l'applet de commande PowerShell Get-InitiatorPort.

```
# Get the IQN for iSCSI
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'iSCSI'} | Select-
Object -Property NodeAddress
```

```
# Get the WWPN for FC
Get-InitiatorPort | Where \{$_.ConnectionType -eq 'Fibre Channel'} |
Select-Object -Property PortAddress
```

```
# While adding initiator to the initiator group in case of FC, make sure
to provide the initiator(PortAddress) in the standard WWPN format
```

L'IQN pour iSCSI sur Windows Server peut également être vérifié dans la configuration des propriétés de l'initiateur iSCSI.

- Créez une LUN à l'aide de l'assistant de création de LUN et associez-la au groupe initiateur créé.

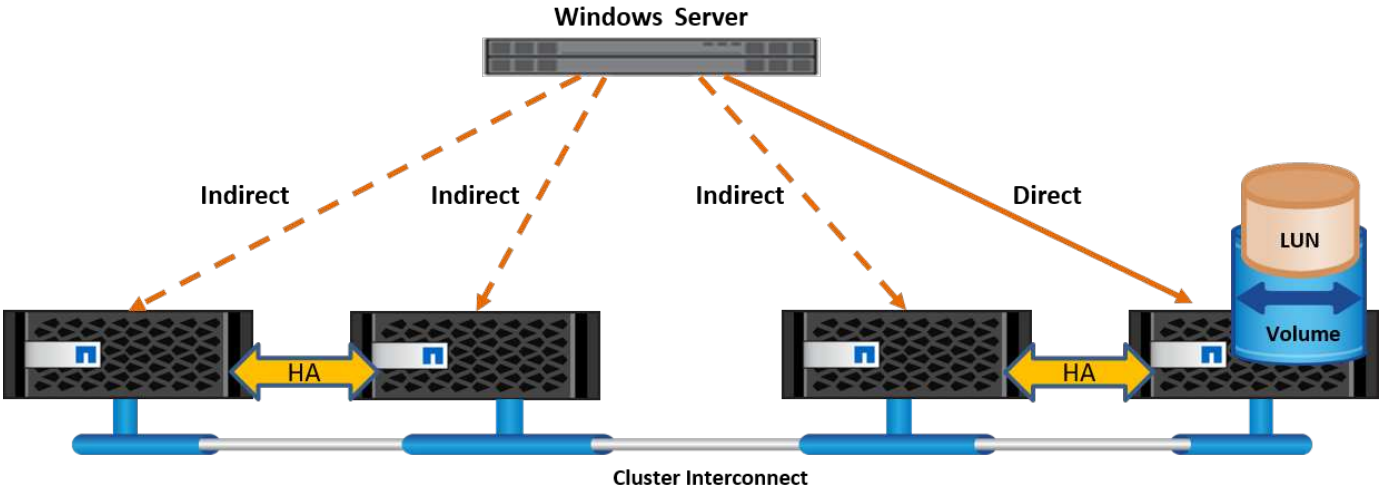
Intégration de l'hôte

Windows Server utilise l'extension ALUA (Asymmetrical Logical Unit Access) MPIO pour déterminer les chemins directs et indirects vers les LUN. Bien que chaque LIF appartenant à un SVM accepte des demandes de lecture/écriture pour ses LUN, seul un des nœuds du cluster possède les disques sur lesquels cette LUN est supports à un moment donné. Cela divise les chemins disponibles pour une LUN en deux types, directs ou indirects, comme indiqué dans la figure suivante.

Un chemin direct pour une LUN est un chemin sur lequel résident les LIFs d'un SVM et la LUN accédée sur le même nœud. Pour passer d'un port cible physique à un disque, il n'est pas nécessaire de traverser

l'interconnexion de cluster.

Les chemins indirects sont des chemins de données sur lesquels les LIFs du SVM et la LUN accédée résident sur différents nœuds. Les données doivent traverser l'interconnexion de cluster pour passer d'un port cible physique à un disque.



MPIO

ONTAP fournit un stockage hautement disponible dans lequel plusieurs chemins d'accès peuvent exister entre le contrôleur de stockage et le serveur Windows. Les chemins d'accès multiples correspondent à la capacité d'avoir plusieurs chemins de données entre un serveur et une baie de stockage. Les chemins d'accès multiples protègent contre les pannes matérielles (coupures de câbles, défaillance du switch et de l'adaptateur de bus hôte [HBA], etc.). Ils peuvent également offrir des limites de performances plus élevées en utilisant les performances d'agrégat de plusieurs connexions. Lorsqu'un chemin ou une connexion devient indisponible, le logiciel de chemins d'accès multiples déplace automatiquement la charge vers l'un des autres chemins disponibles. La fonctionnalité MPIO combine les chemins physiques multiples vers le stockage comme un chemin logique unique utilisé pour l'accès aux données, afin d'assurer la résilience du stockage et l'équilibrage de la charge. Pour utiliser cette fonctionnalité, la fonctionnalité MPIO doit être activée sur Windows Server.

Activez MPIO

Pour activer MPIO sur Windows Server, procédez comme suit :

1. Connectez-vous à Windows Server en tant que membre du groupe d'administrateurs.
2. Démarrez Server Manager.
3. Dans la section gérer, cliquez sur Ajouter des rôles et des fonctions.
4. Dans la page Sélectionner les fonctionnalités, sélectionnez E/S multivoies

Configurer MPIO

Lorsque vous utilisez le protocole iSCSI, vous devez demander à Windows Server d'appliquer la prise en charge des chemins d'accès multiples aux périphériques iSCSI dans les propriétés MPIO.

Pour configurer MPIO sur Windows Server, procédez comme suit :

1. Connectez-vous à Windows Server en tant que membre du groupe d'administrateurs.
2. Démarrez Server Manager.

3. Dans la section Outils, cliquez sur MPIO.
4. Dans Propriétés MPIO sur Discover Multi-Path, sélectionnez Add support for iSCSI Devices et cliquez sur Add. Une invite vous demande ensuite de redémarrer l'ordinateur.
5. Redémarrez Windows Server pour voir le périphérique MPIO répertorié dans la section périphériques MPIO des Propriétés MPIO.

Configurez iSCSI

Pour détecter le stockage en mode bloc iSCSI sur Windows Server, procédez comme suit :

1. Connectez-vous à Windows Server en tant que membre du groupe d'administrateurs.
2. Démarrez Server Manager.
3. Dans la section Outils, cliquez sur initiateur iSCSI.
4. Sous l'onglet découverte, cliquez sur découvrir le portail.
5. Fournir l'adresse IP des LIFs associées au SVM créé pour le protocole NetApp Storage for SAN. Cliquez sur Avancé, configurez les informations dans l'onglet général, puis cliquez sur OK.
6. L'initiateur iSCSI détecte automatiquement la cible iSCSI et la répertorie dans l'onglet cibles.
7. Sélectionnez la cible iSCSI dans cibles découvertes. Cliquez sur connexion pour ouvrir la fenêtre connexion à la cible.
8. Vous devez créer plusieurs sessions à partir de l'hôte Windows Server vers les LIFs iSCSI cibles sur le cluster de stockage NetApp. Pour ce faire, procédez comme suit :
9. Dans la fenêtre se connecter à la cible, sélectionnez Activer MPIO et cliquez sur Avancé.
10. Dans Paramètres avancés sous l'onglet général, sélectionnez la carte locale en tant qu'initiateur Microsoft iSCSI et sélectionnez l'adresse IP de l'initiateur et l'adresse IP du portail cible.
11. Vous devez également vous connecter à l'aide du second chemin. Par conséquent, répétez les étapes 5 à 8, mais cette fois, sélectionnez l'adresse IP de l'initiateur et l'adresse IP du portail cible pour le second chemin.
12. Sélectionnez la cible iSCSI dans cibles découvertes dans la fenêtre principale des propriétés iSCSI et cliquez sur Propriétés.
13. La fenêtre Propriétés indique que plusieurs sessions ont été détectées. Sélectionnez la session, cliquez sur périphériques, puis cliquez sur MPIO pour configurer la stratégie d'équilibrage de charge. Tous les chemins configurés pour le périphérique sont affichés et toutes les stratégies d'équilibrage de charge sont prises en charge. NetApp recommande généralement la permutation circulaire avec sous-ensemble, et ce paramètre est le paramètre par défaut pour les baies pour lesquelles le protocole ALUA est activé. Round Robin est la valeur par défaut pour les baies actives/actives qui ne prennent pas en charge ALUA.

Détecter le stockage bloc

Pour détecter un stockage en mode bloc iSCSI ou FC sur Windows Server, effectuez les opérations suivantes :

1. Cliquez sur gestion de l'ordinateur dans la section Outils du Gestionnaire de serveur.
2. Dans gestion de l'ordinateur, cliquez sur la section gestion des disques dans le stockage, puis cliquez sur autres actions et sur Nouvelle analyse des disques. Les LUN iSCSI brutes s'affichent alors.
3. Cliquez sur la LUN découverte et mettez-la en ligne. Sélectionnez ensuite initialiser le disque à l'aide de la partition MBR ou GPT. Créez un nouveau volume simple en indiquant la taille du volume et la lettre du lecteur et formatez-le à l'aide de FAT, FAT32, NTFS ou du système de fichiers résilient (ReFS).

Et des meilleures pratiques

- NetApp recommande d'activer le provisionnement fin sur les volumes hébergeant les LUN.
- Pour éviter les problèmes de chemins d'accès multiples, NetApp recommande d'utiliser toutes les sessions de 10 Gbits ou toutes les sessions de 1 Gbit vers une LUN donnée.
- NetApp vous recommande de vérifier que le protocole ALUA est activé sur le système de stockage. ALUA est activé par défaut sur ONTAP.
- Sur l'hôte Windows Server auquel est mappée la LUN NetApp, activez le service iSCSI (TCP-in) pour le service entrant et le service iSCSI (TCP-out) pour le service sortant dans les paramètres du pare-feu. Ces paramètres permettent au trafic iSCSI de passer de et vers l'hôte Hyper-V et le contrôleur NetApp.

Provisionnement des LUN NetApp sur le serveur Nano

Prérequis

En plus des conditions préalables mentionnées dans la section précédente, le rôle de stockage doit être activé du côté Nano Server. Par exemple, Nano Server doit être déployé à l'aide de l'option -Storage. Pour déployer Nano Server, reportez-vous à la section ""Déployez Nano Server."«

Déploiement

Pour provisionner des LUN NetApp sur un serveur Nano, procédez comme suit :

1. Connectez-vous au Nano Server à distance en suivant les instructions de la section ""Connectez-vous au Nano Server"."
2. Pour configurer iSCSI, exécutez les applets de commande PowerShell suivantes sur le Nano Server :

```
# Start iSCSI service, if it is not already running
Start-Service msiscsi
```

```
# Create a new iSCSI target portal
New-IscsiTargetPortal -TargetPortalAddress <SVM LIF>
```

```
# View the available iSCSI targets and their node address
Get-IscsiTarget
```

```
# Connect to iSCSI target
Connect-IscsiTarget -NodeAddress <NodeAddress>
```

```
# NodeAddress is retrived in above cmdlet Get-IscsiTarget
# OR
Get-IscsiTarget | Connect-IscsiTarget
```

```
# View the established iSCSI session
Get-IscsiSession
```

```
# Note the InitiatorNodeAddress retrieved in the above cmdlet Get-
IscsiSession. This is the IQN for Nano server and this needs to be added
in the Initiator group on NetApp Storage
```

```
# Rescan the disks
Update-HostStorageCache
```

3. Ajouter un initiateur au groupe initiateur.

```
Add the InitiatorNodeAddress retrieved from the cmdlet Get-IscsiSession
to the Initiator Group on NetApp Controller
```

4. Configurer MPIO.

```
# Enable MPIO Feature
Enable-WindowsOptionalFeature -Online -FeatureName MultipathIo
```

```
# Get the Network adapters and their IPs
Get-NetIPAddress -AddressFamily IPv4 -PrefixOrigin <Dhcp or Manual>
```

```
# Create one MPIO-enabled iSCSI connection per network adapter
Connect-IscsiTarget -NodeAddress <NodeAddress> -IsPersistent $True
-IsMultipathEnabled $True -InitiatorPortalAddress <IP Address of
ethernet adapter>
```

```
# NodeAddress is retrieved from the cmdlet Get-IscsiTarget
# IPs are retrieved in above cmdlet Get-NetIPAddress
```

```
# View the connections
Get-IscsiConnection
```

5. Détecter le stockage bloc.

```
# Rescan disks
Update-HostStorageCache
```

```
# Get details of disks
Get-Disk
```

```
# Initialize disk
Initialize-Disk -Number <DiskNumber> -PartitionStyle <GPT or MBR>
```

```
# DiskNumber is retrived in the above cmdlet Get-Disk
# Bring the disk online
Set-Disk -Number <DiskNumber> -IsOffline $false
```

```
# Create a volume with maximum size and default drive letter
New-Partition -DiskNumber <DiskNumber> -UseMaximumSize
-AssignDriveLetter
```

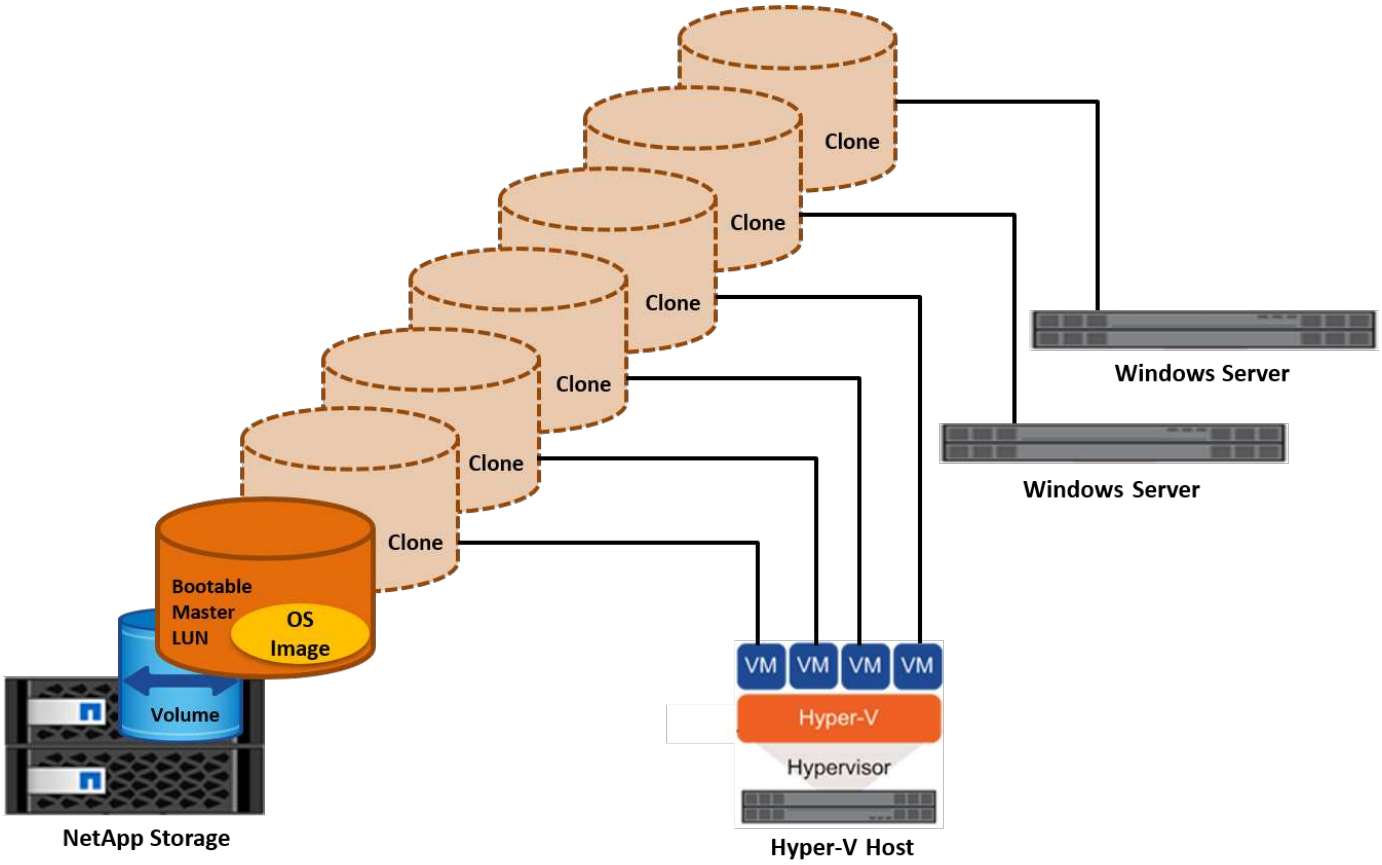
```
# To choose the size and drive letter use -Size and -DriveLetter
parameters
# Format the volume
Format-Volume -DriveLetter <DriveLetter> -FileSystem <FAT32 or NTFS or
REFS>
```

Démarrage à partir du réseau SAN

Un hôte physique (serveur) ou une machine virtuelle Hyper-V peut démarrer le système d'exploitation Windows Server directement à partir d'un LUN NetApp au lieu de son disque dur interne. Dans l'approche de démarrage à partir du SAN, l'image du système d'exploitation à partir de réside sur un LUN NetApp connecté à un hôte physique ou à une machine virtuelle. Dans le cas d'un hôte physique, le HBA de l'hôte physique est configuré pour utiliser le LUN NetApp pour le démarrage. Dans le cas d'une machine virtuelle, le LUN NetApp est connecté en tant que disque pass-through pour le démarrage.

Approche NetApp FlexClone

Grâce à la technologie NetApp FlexClone, les LUN de démarrage avec une image du système d'exploitation peuvent être clonées instantanément et reliées aux serveurs et aux serveurs virtuels pour fournir rapidement des images de système d'exploitation propres, comme illustré dans la figure suivante.



Démarrage à partir du SAN pour l'hôte physique

Prérequis

- L'hôte physique (serveur) dispose d'une carte HBA iSCSI ou FC appropriée.
- Vous avez téléchargé un pilote de périphérique HBA approprié pour le serveur prenant en charge Windows Server.
- Le serveur dispose d'un lecteur de CD/DVD ou d'un support virtuel approprié pour insérer l'image ISO Windows Server et le pilote de périphérique HBA a été téléchargé.
- Une LUN NetApp iSCSI ou FC est provisionnée sur le contrôleur de stockage NetApp.

Déploiement

Pour configurer le démarrage à partir du réseau SAN pour un hôte physique, procédez comme suit :

1. Activez BootBIOS sur le HBA du serveur.
2. Pour les HBA iSCSI, configurez l'adresse IP de l'initiateur, le nom du nœud iSCSI et le mode d'amorçage de l'adaptateur dans les paramètres du BIOS d'amorçage.
3. Lors de la création d'un groupe initiateur pour iSCSI et/ou FC sur un contrôleur de stockage NetApp, ajoutez l'initiateur HBA du serveur au groupe. L'initiateur HBA du serveur est le WWPN correspondant au HBA FC ou au nom du nœud iSCSI du HBA iSCSI.
4. Créez une LUN sur le contrôleur de stockage NetApp avec l'ID de LUN 0 et associez-la au groupe initiateur créé à l'étape précédente. Cette LUN sert de LUN de démarrage.
5. Limitez le HBA à un seul chemin vers la LUN de démarrage. Des chemins supplémentaires peuvent être ajoutés après l'installation de Windows Server sur la LUN de démarrage pour exploiter la fonctionnalité de

chemins d'accès multiples.

6. Utilisez l'utilitaire BootBIOS du HBA pour configurer le LUN en tant que périphérique d'amorçage.
7. Redémarrez l'hôte et accédez à l'utilitaire BIOS de l'hôte.
8. Configurez le BIOS hôte pour que la LUN de démarrage soit le premier périphérique dans l'ordre de démarrage.
9. À partir de l'ISO Windows Server, lancez la configuration de l'installation.
10. Lorsque l'installation vous demande « où voulez-vous installer Windows ? », cliquez sur Charger le pilote en bas de l'écran d'installation pour lancer la page Sélectionner le pilote à installer. Indiquez le chemin du pilote de périphérique HBA téléchargé précédemment et terminez l'installation du pilote.
11. La LUN de démarrage créée précédemment doit maintenant être visible sur la page d'installation de Windows. Sélectionnez la LUN de démarrage pour l'installation de Windows Server sur la LUN de démarrage et terminez l'installation.

Démarrage à partir du SAN pour la machine virtuelle

Pour configurer le démarrage à partir du SAN pour une machine virtuelle, procédez comme suit :

Déploiement

1. Lors de la création d'un groupe initiateur pour iSCSI ou FC sur un contrôleur de stockage NetApp, ajoutez l'IQN pour iSCSI ou le WWN pour FC du serveur Hyper-V au contrôleur.
2. Créez des LUN ou des clones de LUN sur le contrôleur de stockage NetApp et associez-les au groupe initiateur créé à l'étape précédente. Ces LUN servent de LUN de démarrage pour les machines virtuelles.
3. Détecter les LUN sur le serveur Hyper-V, les mettre en ligne et les initialiser.
4. Mettez les LUN hors ligne.
5. Créez des machines virtuelles avec l'option connecter un disque dur virtuel ultérieurement sur la page connecter un disque dur virtuel.
6. Ajout d'une LUN en tant que disque pass-through à une VM
   a. Ouvrez les paramètres de la machine virtuelle.
   b. Cliquez sur contrôleur IDE 0, sélectionnez disque dur, puis cliquez sur Ajouter. Si vous sélectionnez IDE Controller 0, ce disque devient le premier périphérique d'amorçage pour la machine virtuelle.
   c. Sélectionnez disque dur physique dans les options disque dur et sélectionnez un disque dans la liste comme disque intermédiaire. Les disques sont les LUN configurés dans les étapes précédentes.
7. Installez Windows Server sur le disque d'intercommunication.

Et des meilleures pratiques

- Assurez-vous que les LUN sont hors ligne. Sinon, le disque ne peut pas être ajouté en tant que disque pass-through à une machine virtuelle.
- Lorsqu'il existe plusieurs LUN, veillez à noter le numéro de disque de la LUN dans la gestion de disque. Cette opération est nécessaire car les disques répertoriés pour la machine virtuelle sont répertoriés avec le numéro de disque. De même, la sélection du disque en tant que disque pass-through pour la machine virtuelle est basée sur ce numéro de disque.
- NetApp recommande d'éviter le regroupement de cartes réseau pour les cartes réseau iSCSI.
- NetApp recommande d'utiliser le MPIO ONTAP configuré sur l'hôte à des fins de stockage.

Le provisionnement dans les environnements SMB

ONTAP fournit un stockage NAS résilient et hautes performances pour les machines virtuelles Hyper-V utilisant le protocole SMB3.

Lorsqu'un SVM est créé avec le protocole CIFS, un serveur CIFS s'exécute en plus du SVM faisant partie du domaine Windows Active Directory. Les partages SMB peuvent être utilisés pour un répertoire local et pour héberger les charges de travail Hyper-V et SQL Server. ONTAP prend en charge les fonctionnalités SMB 3.0 suivantes :

- Pointeurs permanents (partages de fichiers disponibles en continu)
- Protocole témoin
- Basculement client en cluster
- Sensibilisation au scale-out
- ODX
- VSS distant

Provisionnement des partages SMB sur Windows Server

Prérequis

L'utilisation d'un stockage NetApp dans des environnements NAS sous Windows Server présente les conditions suivantes :

- Le cluster ONTAP dispose d'une licence CIFS valide.
- Au moins un agrégat est créé.
- Une interface logique de données (LIF) est créée et la LIF de données doit être configurée pour CIFS.
- Un serveur de domaine Windows Active Directory configuré par DNS et des informations d'identification d'administrateur de domaine sont présentes.
- Chaque nœud du cluster NetApp est synchronisé avec le contrôleur de domaine Windows.

Contrôleur de domaine Active Directory

Il est possible de joindre un contrôleur de stockage NetApp à un serveur Active Directory similaire à un serveur Windows et de le faire fonctionner au sein de celui-ci. Lors de la création du SVM, vous pouvez configurer le DNS en fournissant le nom de domaine et les détails du serveur de noms. Le SVM tente de rechercher un contrôleur de domaine Active Directory en interrogeant le DNS pour un serveur LDAP (Active Directory/Lightweight Directory Access Protocol) d'une manière similaire à Windows Server.

Pour que la configuration CIFS fonctionne correctement, les contrôleurs de stockage NetApp doivent être synchronisés dans le temps avec le contrôleur de domaine Windows. NetApp recommande de ne pas dépasser cinq minutes entre le contrôleur de domaine Windows et le contrôleur de stockage NetApp. Il est recommandé de configurer le serveur NTP (Network Time Protocol) afin que le cluster ONTAP se synchronise avec une source de temps externe. Pour configurer le contrôleur de domaine Windows en tant que serveur NTP, exécutez la commande suivante sur votre cluster ONTAP :

```
$domainControllerIP = "<input IP Address of windows domain controller>"
cluster::> system services ntp server create -server $domainControllerIP
```

Déploiement

1. Créer un SVM avec le protocole NAS CIFS activé Il est possible de créer un SVM avec l'une des méthodes suivantes :
   - Commandes CLI sur ONTAP
   - System Manager
   - Kit NetApp PowerShell
2. Configuration du protocole CIFS
   a. Indiquez le nom du serveur CIFS.
   b. Indiquez l'Active Directory auquel le serveur CIFS doit être joint. Vous devez disposer des informations d'identification de l'administrateur de domaine pour joindre le serveur CIFS à Active Directory.
3. Assigner le SVM avec des LIFs sur chaque nœud de cluster.
4. Démarrer le service CIFS sur le SVM
5. Créez un volume avec le style de sécurité NTFS à partir de l'agrégat.
6. Créer un qtree sur le volume (facultatif).
7. Créez des partages correspondant au volume ou au répertoire qtree afin qu'ils soient accessibles depuis Windows Server. Sélectionnez Activer la disponibilité continue pour Hyper-V lors de la création du partage si celui-ci est utilisé pour le stockage Hyper-V. Cela permet une haute disponibilité pour les partages de fichiers.
8. Modifiez le partage créé et modifiez les autorisations nécessaires pour accéder au partage. Les autorisations du partage SMB doivent être configurées pour accorder l'accès aux comptes d'ordinateur de tous les serveurs accédant à ce partage.

Intégration de l'hôte

Le protocole NAS CIFS est intégré en mode natif à ONTAP. Par conséquent, Windows Server n'a pas besoin d'un logiciel client supplémentaire pour accéder aux données sur ONTAP. Un contrôleur de stockage NetApp apparaît sur le réseau en tant que serveur de fichiers natif et prend en charge l'authentification Microsoft Active Directory.

Pour détecter le partage CIFS créé précédemment avec Windows Server, procédez comme suit :

1. Connectez-vous à Windows Server en tant que membre du groupe d'administrateurs.
2. Accédez à run.exe et saisissez le chemin d'accès complet du partage CIFS créé pour accéder au partage.
3. Pour mapper le partage de façon permanente sur le serveur Windows, cliquez avec le bouton droit de la souris sur ce PC, cliquez sur connecter un lecteur réseau et indiquez le chemin du partage CIFS.
4. Certaines tâches de gestion CIFS peuvent être effectuées à l'aide de la console MMC (Microsoft Management Console). Avant d'effectuer ces tâches, vous devez connecter la console MMC au stockage ONTAP à l'aide des commandes de menu MMC.
   a. Pour ouvrir la console MMC dans Windows Server, cliquez sur gestion de l'ordinateur dans la section Outils de Server Manager.
   b. Cliquez sur autres actions et connectez-vous à un autre ordinateur, ce qui ouvre la boîte de dialogue Sélectionner un ordinateur.
   c. Entrer le nom du serveur CIFS ou l'adresse IP du LIF du SVM pour se connecter au serveur CIFS.
   d. Développez Outils système et dossiers partagés pour afficher et gérer les fichiers, sessions et partages ouverts.

Et des meilleures pratiques

- Pour vérifier qu'il n'y a pas de temps d'indisponibilité lorsqu'un volume est déplacé d'un nœud vers un autre ou en cas de défaillance d'un nœud, NetApp vous recommande d'activer l'option de disponibilité continue sur le partage de fichiers.
- Lors du provisionnement d'ordinateurs virtuels dans un environnement Hyper-V-over-SMB, NetApp vous recommande d'activer la fonction de déchargement des copies sur le système de stockage. Le temps de provisionnement des ordinateurs virtuels est ainsi réduit.
- Si le cluster de stockage héberge plusieurs charges de travail SMB, telles que SQL Server, Hyper-V et des serveurs CIFS, NetApp recommande d'héberger différentes charges de travail SMB sur des SVM distincts, sur des agrégats distincts. Cette configuration est avantageuse, car chacune de ces charges de travail garantit une disposition unique du réseau et des volumes de stockage.
- NetApp recommande de connecter les hôtes Hyper-V et le stockage ONTAP à un réseau de 10 Go, le cas échéant. Dans le cas d'une connectivité réseau de 1 Go, NetApp recommande de créer un groupe d'interfaces composé de plusieurs ports de 1 Go.
- Lors de la migration de machines virtuelles d'un partage SMB 3.0 vers un autre, NetApp recommande d'activer la fonctionnalité de déchargement des copies CIFS sur le système de stockage afin d'accélérer la migration.

Choses à retenir

- Lorsque vous provisionnez des volumes pour les environnements SMB, les volumes doivent être créés avec le style de sécurité NTFS.
- Les paramètres de temps des nœuds du cluster doivent être configurés en conséquence. Utilisez le protocole NTP si le serveur CIFS NetApp doit participer au domaine Windows Active Directory.
- Les pointeurs permanents fonctionnent uniquement entre les nœuds d'une paire haute disponibilité.
- Le protocole témoin fonctionne uniquement entre les nœuds d'une paire haute disponibilité.
- Les partages de fichiers disponibles en continu sont pris en charge uniquement pour les charges de travail Hyper-V et SQL Server.
- Le multicanal SMB est pris en charge à partir de ONTAP 9.4.
- RDMA n'est pas pris en charge.
- Les références ne sont pas prises en charge.

Provisionnement des partages SMB sur Nano Server

Le serveur nano n'a pas besoin d'un logiciel client supplémentaire pour accéder aux données du partage CIFS sur un contrôleur de stockage NetApp.

Pour copier des fichiers de Nano Server vers un partage CIFS, exécutez les applets de commande suivantes sur le serveur distant :

```
$ip = "<input IP Address of the Nano Server>"
```

```
# Create a New PS Session to the Nano Server
$session = New-PSSession -ComputerName $ip -Credential ~\Administrator
```

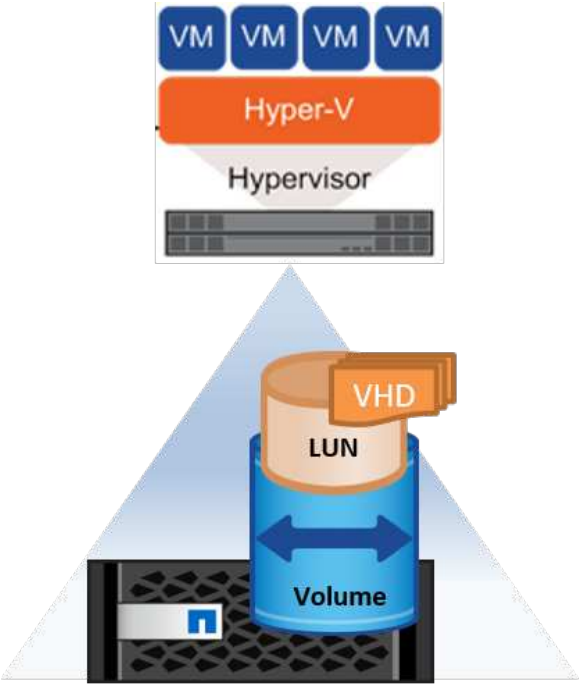
```
 Copy-Item -FromSession $s -Path C:\Windows\Logs\DISM\dism.log
-Destination \\cifsshare
* `cifsshare` Est le partage CIFS sur le contrôleur de stockage NetApp.
* Pour copier des fichiers sur Nano Server, exécutez l'applet de commande
suivante :
```

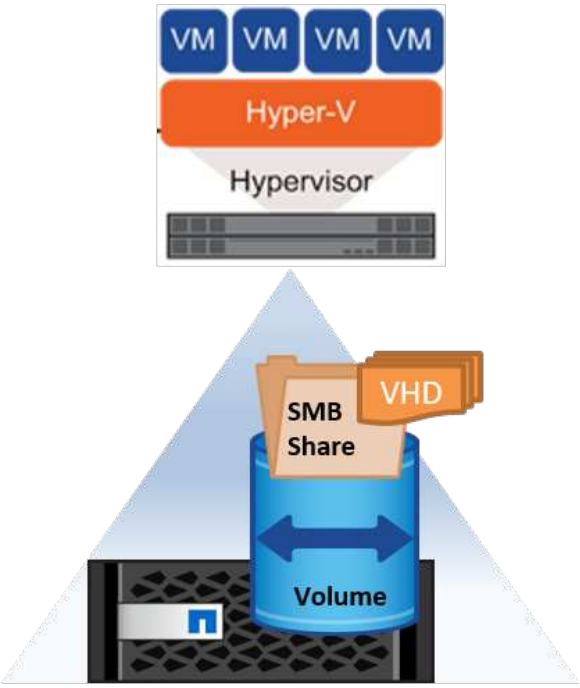
+
Copy-Item -ToSession $s -Path \\cifsshare\<file> -Destination C:\

Pour copier l'intégralité du contenu d'un dossier, spécifiez le nom du dossier et utilisez le paramètre -RECURSE à la fin de l'applet de commande.

Infrastructure de stockage Hyper-V sur NetApp

Une infrastructure de stockage Hyper-V peut être hébergée sur des systèmes de stockage ONTAP. Le stockage d'Hyper-V pour stocker les fichiers de la machine virtuelle et ses disques peut être fourni à l'aide de LUN NetApp ou de partages CIFS NetApp, comme illustré dans la figure ci-dessous.





Stockage Hyper-V sur LUN NetApp

- Provisionner un LUN NetApp sur le serveur Hyper-V. Pour plus d'informations, reportez-vous à la section «"Le provisionnement dans des environnements SAN"."
- Ouvrez Hyper-V Manager dans la section Outils de Server Manager.
- Sélectionnez le serveur Hyper-V et cliquez sur Paramètres Hyper-V.
- Spécifiez le dossier par défaut pour stocker la machine virtuelle et son disque en tant que LUN. Le chemin

par défaut est alors défini comme LUN pour le stockage Hyper-V. Si vous souhaitez spécifier explicitement le chemin d'accès à une machine virtuelle, vous pouvez le faire lors de la création de la machine virtuelle.

Stockage Hyper-V sur NetApp CIFS

Avant de commencer les étapes énumérées dans cette section, passez en revue la section «"Le provisionnement dans les environnements SMB"." Pour configurer le stockage Hyper-V sur le partage CIFS NetApp, effectuez les opérations suivantes :

1. Ouvrez Hyper-V Manager dans la section Outils de Server Manager.
2. Sélectionnez le serveur Hyper-V et cliquez sur Paramètres Hyper-V.
3. Spécifiez le dossier par défaut pour stocker la machine virtuelle et son disque en tant que partage CIFS. Le chemin par défaut est alors défini comme partage CIFS pour le stockage Hyper-V. Si vous souhaitez spécifier explicitement le chemin d'accès à une machine virtuelle, vous pouvez le faire lors de la création de la machine virtuelle.

Chaque machine virtuelle d'Hyper-V peut à son tour être fournie avec les LUN NetApp et les partages CIFS fournis à l'hôte physique. Cette procédure est la même que pour tout hôte physique. Les méthodes suivantes peuvent être utilisées pour provisionner du stockage sur une VM :

- Ajout d'une LUN de stockage à l'aide de l'initiateur FC au sein de la machine virtuelle
- Ajout d'une LUN de stockage à l'aide de l'initiateur iSCSI dans la machine virtuelle
- Ajout d'un disque physique pass-through à une machine virtuelle
- Ajout de VHD/VHDX à une machine virtuelle à partir de l'hôte

Et des meilleures pratiques

- Lorsqu'un serveur virtuel et ses données sont stockés sur un système de stockage NetApp, NetApp recommande d'exécuter régulièrement la déduplication NetApp au niveau du volume. Cette pratique permet de réaliser d'importantes économies d'espace lorsque des machines virtuelles identiques sont hébergées sur un partage CSV ou SMB. La déduplication s'exécute sur le contrôleur de stockage et n'affecte pas le système hôte ni les performances des machines virtuelles.
- Lorsque vous utilisez des LUN iSCSI pour Hyper-V, assurez-vous d'activer le service iSCSI (TCP-in) pour le trafic entrant et le service iSCSI (TCP-out) pour le trafic sortant dans les paramètres du pare-feu sur l'hôte Hyper-V. Le trafic iSCSI peut ainsi passer de et vers l'hôte Hyper-V et le contrôleur NetApp.
- NetApp recommande de décocher l'option Autoriser le système d'exploitation de gestion à partager cette carte réseau pour le commutateur virtuel Hyper-V. Cela crée un réseau dédié pour les machines virtuelles.

Choses à retenir

- Le provisionnement d'une machine virtuelle à l'aide de la technologie Fibre Channel virtuelle nécessite une carte HBA FC compatible avec la virtualisation N_Port ID. Quatre ports FC au maximum sont pris en charge.
- Si le système hôte est configuré avec plusieurs ports FC et présenté à la machine virtuelle, MPIO doit être installé dans la machine virtuelle pour activer les chemins d'accès multiples.
- Les disques directs ne peuvent pas être provisionnés vers l'hôte si MPIO est utilisé sur cet hôte, car les disques directs ne prennent pas en charge MPIO.
- Le disque utilisé pour les fichiers VHD/VHDx doit utiliser un formatage de 64 Ko pour l'allocation.

Lecture ultérieure

- Pour plus d'informations sur les HBA FC, reportez-vous au "Matrice d'interopérabilité (IMT)" .
- Pour plus d'informations sur la technologie Fibre Channel virtuelle, consultez le document Microsoft "Présentation de Hyper-V Virtual Fibre Channel" page.

Transfert de données allégé

Microsoft ODX, également appelé allègement de la charge des copies, permet des transferts directs de données au sein d'un dispositif de stockage ou entre des dispositifs de stockage compatibles sans transférer les données via l'ordinateur hôte. ONTAP prend en charge la fonction ODX pour les protocoles CIFS et SAN. ODX peut améliorer les performances si les copies se trouvent dans le même volume, réduire l'utilisation du processeur et de la mémoire sur le client et réduire l'utilisation de la bande passante des E/S réseau.

Avec ODX, il est plus rapide et efficace de copier des fichiers au sein des partages SMB, au sein des LUN et entre les partages SMB et les LUN s'ils se trouvent dans le même volume. Cette approche s'avère plus utile dans le cas où plusieurs copies de l'image de référence d'un système d'exploitation (VHD/VHDX) sont requises au sein du même volume. Plusieurs copies de la même image de référence peuvent être réalisées en beaucoup moins de temps si les copies se trouvent dans le même volume. ODX est également appliqué à la migration dynamique du stockage Hyper-V pour le déplacement du stockage des machines virtuelles.

Si la copie se trouve sur plusieurs volumes, les performances peuvent ne pas être nettement supérieures à celles des copies basées sur l'hôte.

Pour activer la fonction ODX sur CIFS, exécutez les commandes CLI suivantes sur le contrôleur de stockage NetApp :

1. Activez ODX pour CIFS.
   #définissez le niveau de privilège sur diagnostic
   cluster:> diagnostic set -privilege

   ```
   #enable the odx feature
   cluster::> vserver cifs options modify -vserver <vserver_name> -copy
   -offload-enabled true
   ```

   ```
   #return to admin privilege level
   cluster::> set privilege admin
   ```

2. Pour activer la fonction ODX sur SAN, exécutez les commandes CLI suivantes sur le contrôleur de stockage NetApp :
   #définissez le niveau de privilège sur diagnostic
   cluster:> diagnostic set -privilege

   ```
   #enable the odx feature
   cluster::> copy-offload modify -vserver <vserver_name> -scsi enabled
   ```

```
#return to admin privilege level
cluster::> set privilege admin
```

Choses à retenir

- Pour CIFS, ODX est disponible uniquement lorsque le client et le serveur de stockage prennent en charge SMB 3.0 et la fonction ODX.
- Pour les environnements SAN, ODX est disponible uniquement lorsque le client et le serveur de stockage prennent en charge la fonctionnalité ODX.

Lecture ultérieure

Pour plus d'informations sur ODX, voir "Amélioration des performances de Microsoft Remote Copy" et "Transferts de données allégés par Microsoft" .

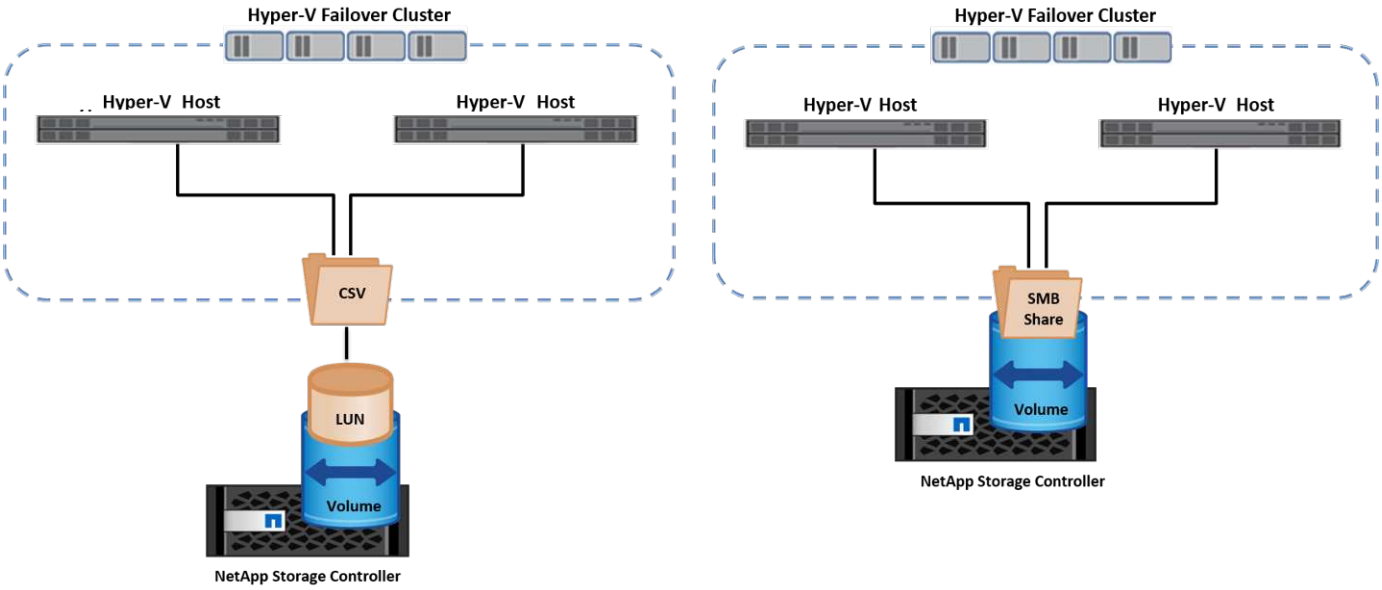
Mise en cluster Hyper-V : haute disponibilité et évolutivité pour les machines virtuelles

Les clusters de basculement offrent une haute disponibilité et une évolutivité aux serveurs Hyper-V. Un cluster de basculement est un groupe de serveurs Hyper-V indépendants qui fonctionnent ensemble pour améliorer la disponibilité et l'évolutivité des machines virtuelles.

Les serveurs en cluster Hyper-V (appelés nœuds) sont connectés par le réseau physique et par un logiciel de cluster. Ces nœuds utilisent un stockage partagé pour stocker les fichiers de la machine virtuelle, notamment les fichiers de configuration, les fichiers des disques durs virtuels (VHD) et les snapshots. Le stockage partagé peut être un partage SMB/CIFS NetApp ou un fichier CSV sur un LUN NetApp, comme illustré ci-dessous. Ce stockage partagé fournit un namespace cohérent et distribué auquel tous les nœuds du cluster peuvent accéder simultanément. Par conséquent, si un nœud tombe en panne dans le cluster, l'autre nœud assure le service par un processus appelé basculement. Les clusters de basculement peuvent être gérés à l'aide du composant logiciel enfichable Failover Cluster Manager et des applets de commande de mise en cluster de basculement Windows PowerShell.

Volumes partagés de cluster

Les CSV permettent à plusieurs nœuds d'un cluster de basculement de disposer simultanément d'un accès en lecture/écriture vers le même LUN NetApp provisionné en tant que volume NTFS ou ReFS. Avec les CSV, les rôles en cluster peuvent basculer rapidement d'un nœud à un autre sans nécessiter de changement de propriétaire de disque, ni de démontage/remontage d'un volume. Les CSV simplifient également la gestion d'un nombre potentiellement important de LUN dans un cluster de basculement. Les CSV proposent un système de fichiers en cluster à usage général qui se superpose au-dessus de NTFS ou ReFS.



Et des meilleures pratiques

- NetApp recommande de désactiver les communications de cluster sur le réseau iSCSI pour empêcher les communications de cluster internes et le trafic CSV de circuler sur le même réseau.
- NetApp recommande de disposer de chemins réseau redondants (plusieurs commutateurs) pour assurer la résilience et la qualité de service.

Choses à retenir

- Les disques utilisés pour CSV doivent être partitionnés avec NTFS ou ReFS. Les disques formatés avec FAT ou FAT32 ne peuvent pas être utilisés pour un CSV.
- Les disques utilisés pour les CSV doivent utiliser un formatage de 64 Ko pour l'allocation.

Lecture ultérieure

Pour plus d'informations sur le déploiement d'un cluster Hyper-V, reportez-vous à l'Annexe B : "Déployez le cluster Hyper-V.".

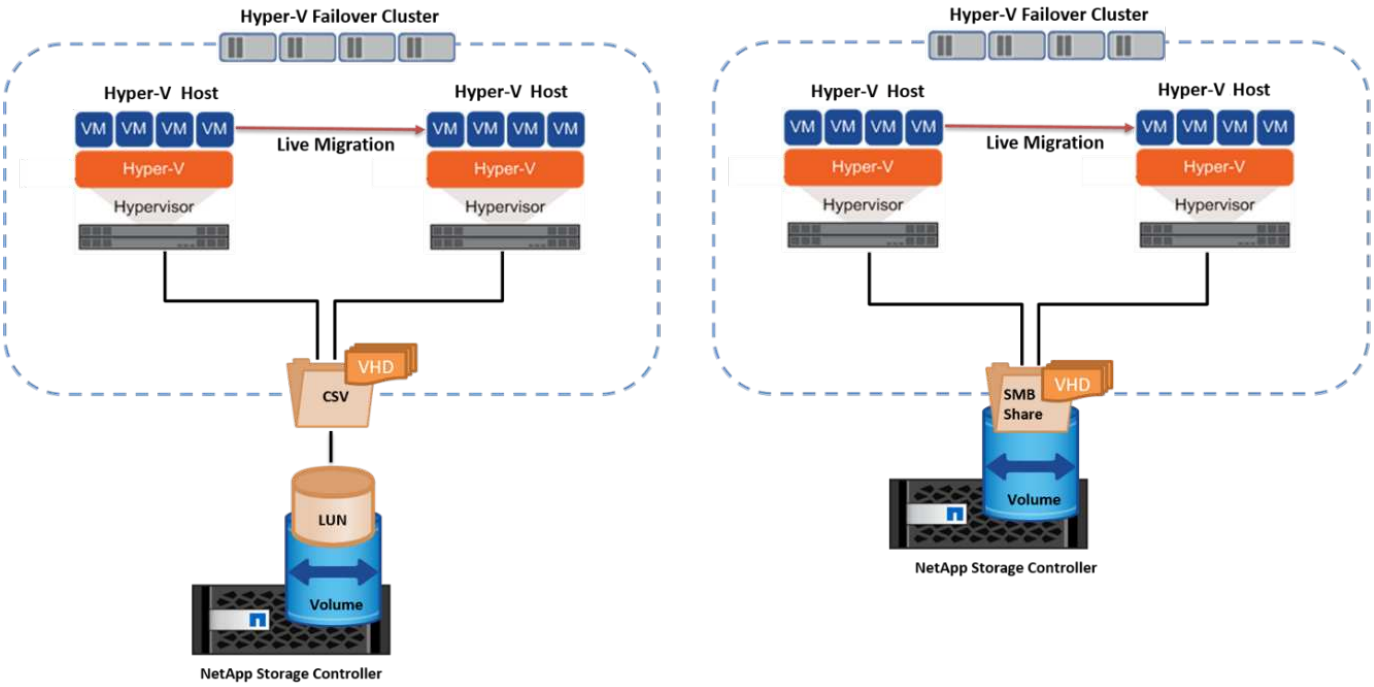
Hyper-V Live migration : migration de machines virtuelles

Il est parfois nécessaire pendant toute la durée de vie des machines virtuelles de les déplacer vers un autre hôte du cluster Windows. Cela peut être nécessaire si l'hôte manque de ressources système ou si l'hôte doit redémarrer pour des raisons de maintenance. De même, il peut être nécessaire de déplacer une machine virtuelle vers une autre LUN ou un autre partage SMB. Cette condition peut être nécessaire si l'espace du LUN ou du partage actuel est insuffisant ou présente des performances inférieures à la valeur attendue. La migration dynamique Hyper-V déplace les machines virtuelles en cours d'exécution d'un serveur Hyper-V physique vers un autre sans affecter la disponibilité des machines virtuelles pour les utilisateurs. Vous pouvez migrer en direct des machines virtuelles entre des serveurs Hyper-V faisant partie d'un cluster de basculement ou entre des serveurs Hyper-V indépendants qui ne font pas partie d'un cluster.

Migration dynamique dans un environnement en cluster

Les machines virtuelles peuvent être déplacées de manière transparente entre les nœuds d'un cluster. La migration des machines virtuelles est instantanée, car tous les nœuds du cluster partagent le même stockage et ont accès à la machine virtuelle et à son disque. La figure suivante illustre la migration en direct dans un

environnement en cluster.



Et des meilleures pratiques

- Disposer d'un port dédié pour le trafic de migration en direct.
- Disposer d'un réseau dédié de migration dynamique des hôtes pour éviter les problèmes liés au réseau pendant la migration.

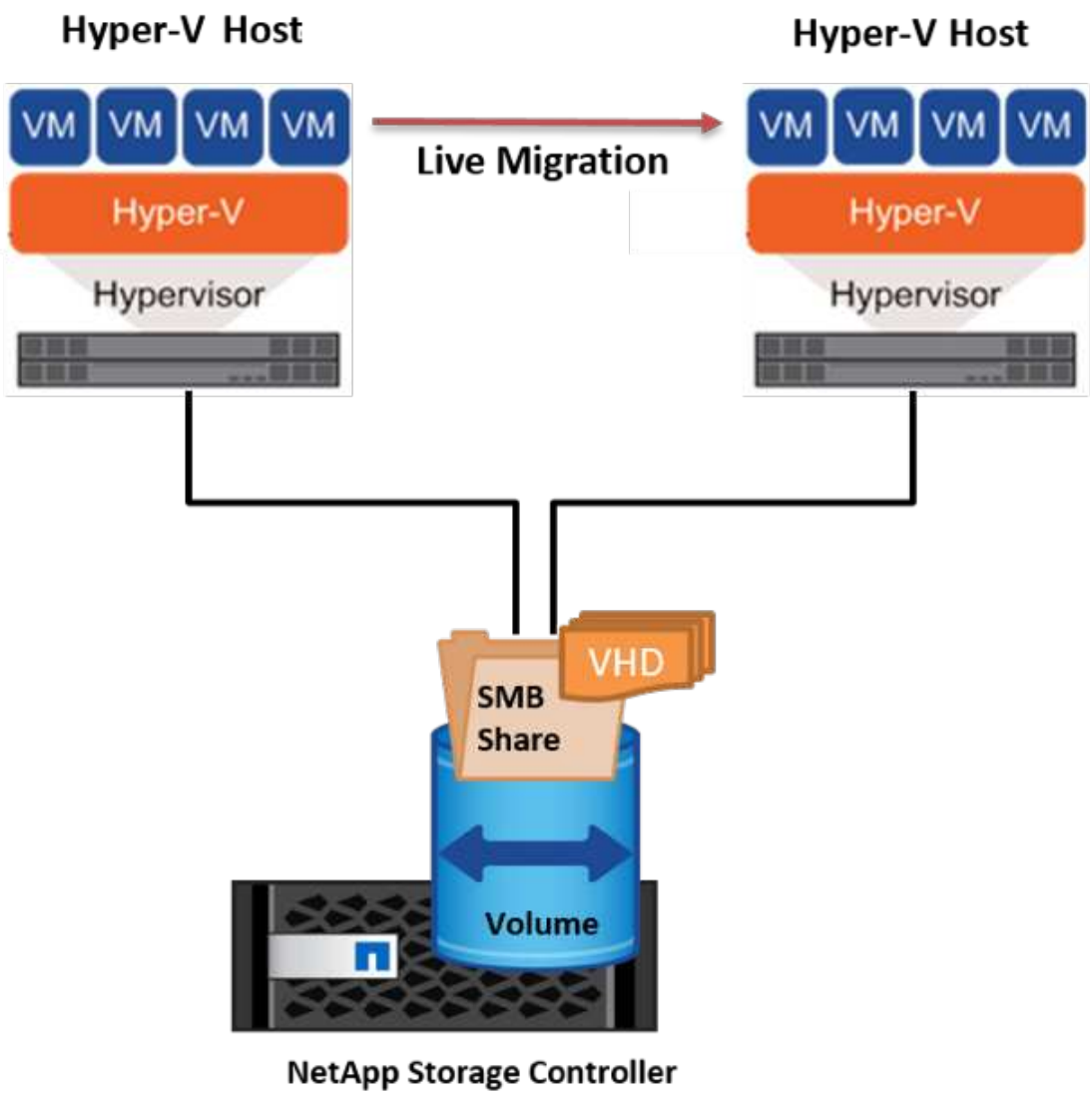
Lecture ultérieure

Pour plus d'informations sur le déploiement de la migration dynamique dans un environnement en cluster, reportez-vous à la section "Annexe C : déploiement de la migration dynamique Hyper-V dans un environnement en cluster".

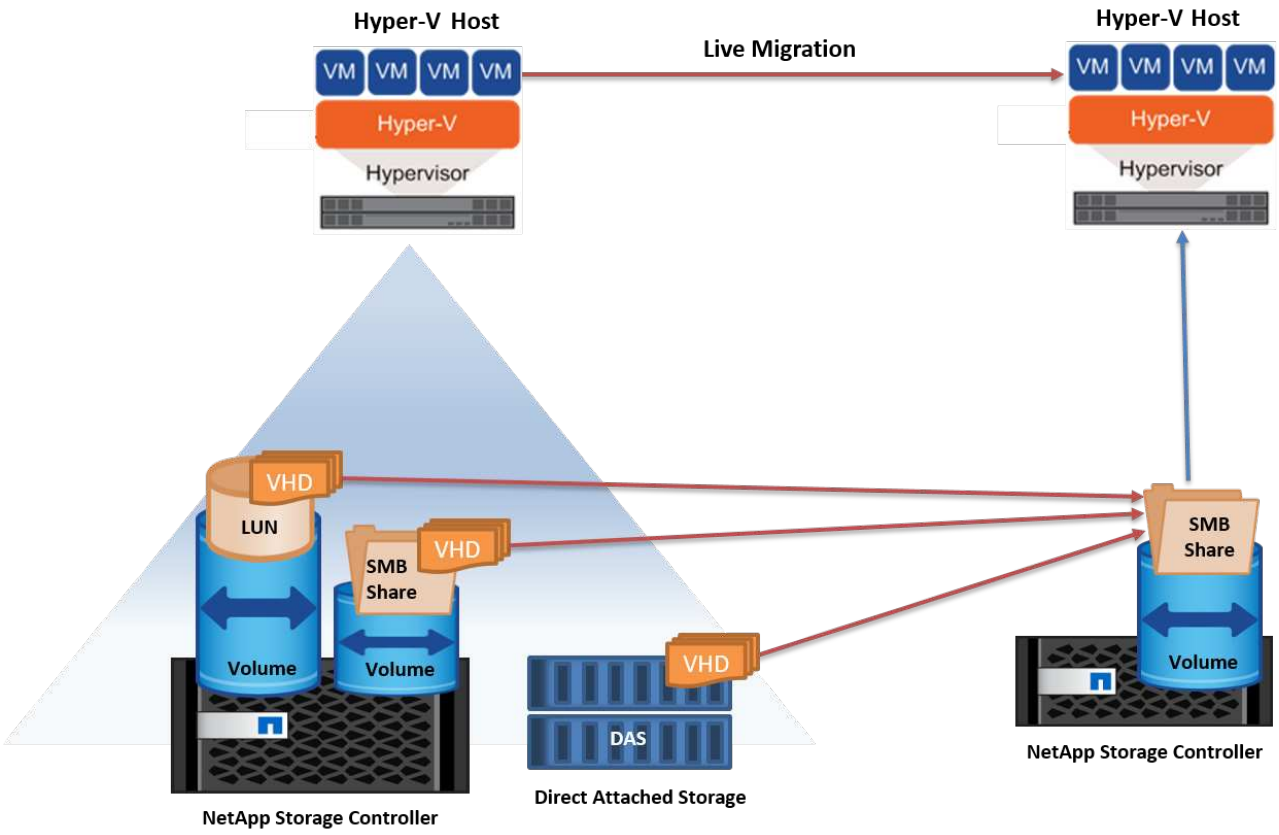
Migration dynamique en dehors d'un environnement en cluster

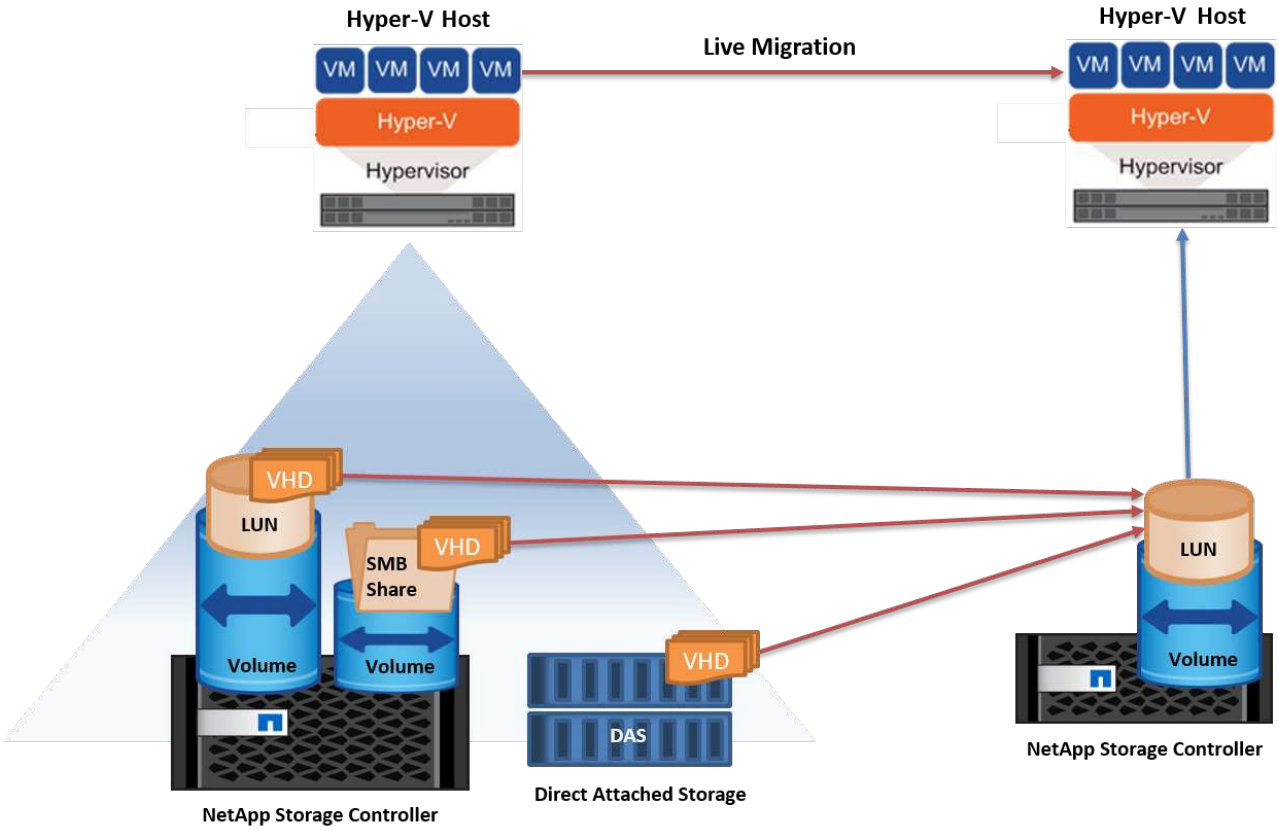
Il est possible de migrer un serveur virtuel en direct entre deux serveurs Hyper-V indépendants non mis en cluster. Ce processus peut utiliser une migration dynamique sans partage ou partagée.

- Dans la migration dynamique partagée, la machine virtuelle est stockée sur un partage SMB. Par conséquent, lorsque vous migrez une machine virtuelle en direct, le stockage de la machine virtuelle reste sur le partage SMB central pour un accès instantané par l'autre nœud, comme illustré ci-dessous.



- Dans le cas d'une migration dynamique sans partage, chaque serveur Hyper-V dispose de son propre stockage local (il peut s'agir d'un partage SMB, d'une LUN ou d'un DAS) et le stockage de la machine virtuelle est local sur son serveur Hyper-V. Lors de la migration en direct d'une machine virtuelle, le stockage de la machine virtuelle est mis en miroir sur le serveur de destination via le réseau client, puis la machine virtuelle est migrée. La machine virtuelle stockée sur le DAS, une LUN ou un partage SMB/CIFS peut être déplacée vers un partage SMB/CIFS sur l'autre serveur Hyper-V, comme illustré dans la figure ci-dessous. Il est également possible de le déplacer vers une LUN, comme illustré dans la seconde figure.





Lecture ultérieure

Pour plus d'informations sur le déploiement de la migration dynamique en dehors d'un environnement en

cluster, reportez-vous à la section "Annexe D : déploiement de la migration dynamique Hyper-V en dehors d'un environnement en cluster".

Hyper-V Storage Live migration

Au cours de la durée de vie d'un serveur virtuel, vous devrez peut-être déplacer le stockage du serveur virtuel (VHD/VHDX) vers un autre LUN ou partage SMB. Cette condition peut être nécessaire si l'espace du LUN ou du partage actuel est insuffisant ou présente des performances inférieures à la valeur attendue.

La LUN ou le partage qui héberge actuellement la machine virtuelle peut être à court d'espace, reconverti ou offre des performances réduites. Dans ces circonstances, la machine virtuelle peut être déplacée sans interruption vers une autre LUN ou un autre partage sur un autre volume, agrégat ou cluster. Ce processus est plus rapide si le système de stockage dispose de fonctionnalités de copie auxiliaire. Les systèmes de stockage NetApp sont dotés de la fonctionnalité de copie auxiliaire activée par défaut dans les environnements CIFS et SAN.

La fonctionnalité ODX effectue des copies de fichiers complets ou de sous-fichiers entre deux répertoires résidant sur des serveurs distants. Une copie est créée en copiant les données entre les serveurs (ou le même serveur si les fichiers source et de destination se trouvent tous deux sur le même serveur). La copie est créée sans que le client ait lu les données à partir de la source ou écrit dans la destination. Ce processus réduit l'utilisation du processeur et de la mémoire pour le client ou le serveur et réduit la bande passante E/S du réseau. La copie est plus rapide si elle se trouve dans le même volume. Si la copie se trouve sur plusieurs volumes, les performances peuvent ne pas être nettement supérieures à celles des copies basées sur l'hôte. Avant de procéder à une opération de copie sur l'hôte, vérifiez que les paramètres de déchargement de copie sont configurés sur le système de stockage.

Lorsque la migration dynamique du stockage de machine virtuelle est initiée à partir d'un hôte, la source et la destination sont identifiées, et l'activité de copie est déchargée sur le système de stockage. Étant donné que l'activité est effectuée par le système de stockage, l'utilisation du processeur, de la mémoire ou du réseau de l'hôte est négligeable.

Les contrôleurs de stockage NetApp prennent en charge les différents scénarios d'ODX suivants :

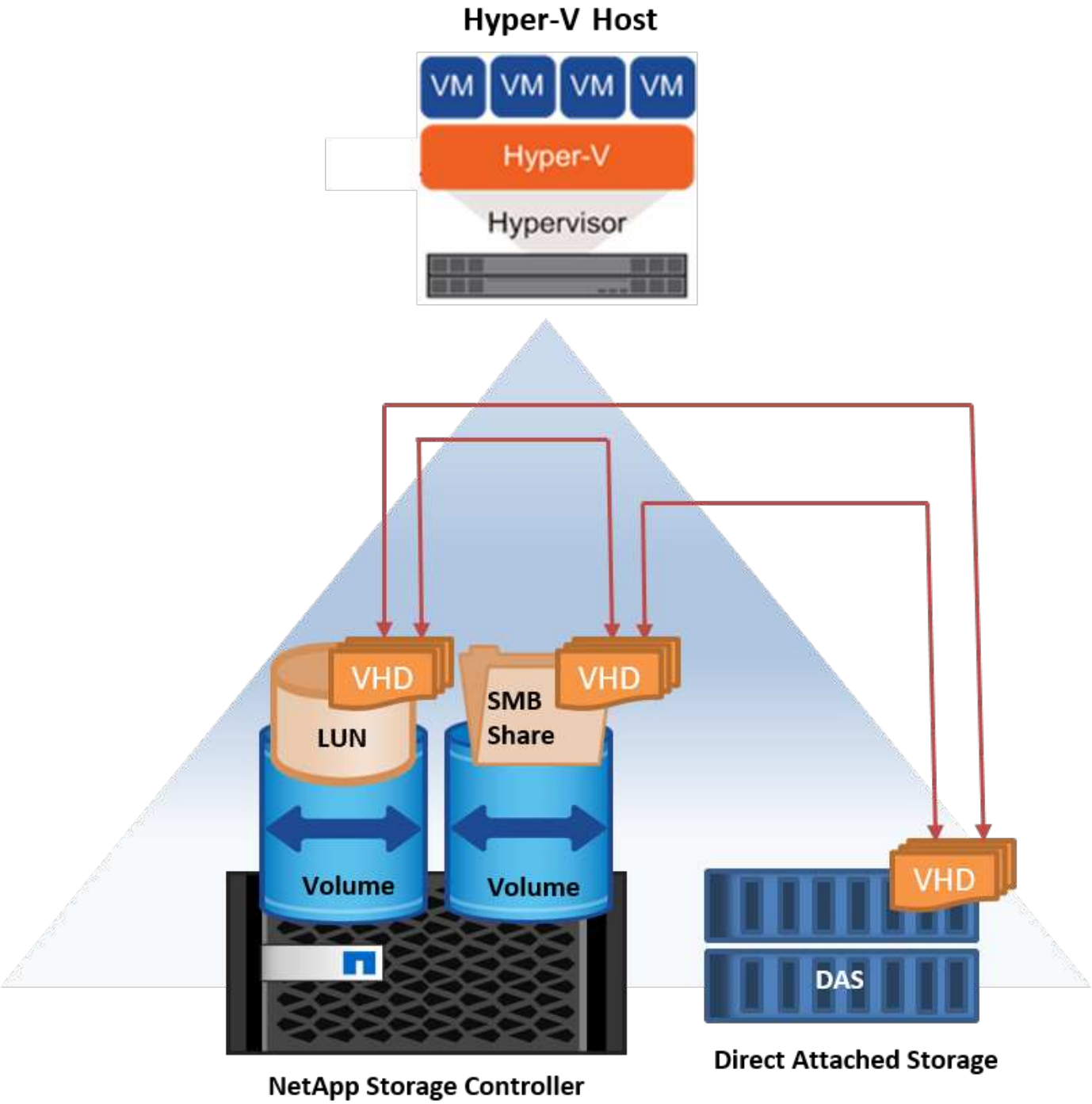
- **IntraSVM.** les données sont détenues par le même SVM :
- **Intravope, intranode.** les fichiers source et de destination ou les LUN résident dans le même volume. La copie s'effectue à l'aide de la technologie de fichiers FlexClone, ce qui offre d'autres avantages en termes de performances de copie à distance.
- **Intervolue, intranode.** les fichiers source et de destination ou les LUN se trouvent sur des volumes différents qui se trouvent sur le même nœud.
- **Intervolue, internœuds.** les fichiers source et de destination ou les LUN se trouvent sur des volumes différents situés sur des nœuds différents.
- **InterSVM.** les données appartiennent à différents SVM.
- **Intervolue, intranode.** les fichiers source et de destination ou les LUN se trouvent sur des volumes différents qui se trouvent sur le même nœud.
- **Intervolue, internœuds.** les fichiers source et de destination ou les LUN se trouvent sur des volumes différents qui se trouvent sur des nœuds différents.
- **Intercluster.** depuis ONTAP 9.0, ODX est également pris en charge pour les transferts de LUN intercluster dans des environnements SAN. ODX intercluster est pris en charge pour les protocoles SAN uniquement, et non pour SMB.

Une fois la migration terminée, les règles de sauvegarde et de réplication doivent être reconfigurées pour refléter le nouveau volume contenant les machines virtuelles. Les sauvegardes précédentes qui ont été

effectuées ne peuvent pas être utilisées.

Le stockage des serveurs virtuels (VHD/VHDX) peut être migré entre les types de stockage suivants :

- Le stockage DAS et le partage SMB
- DAS et LUN
- Un partage SMB et un LUN
- Entre LUN
- Entre partages SMB

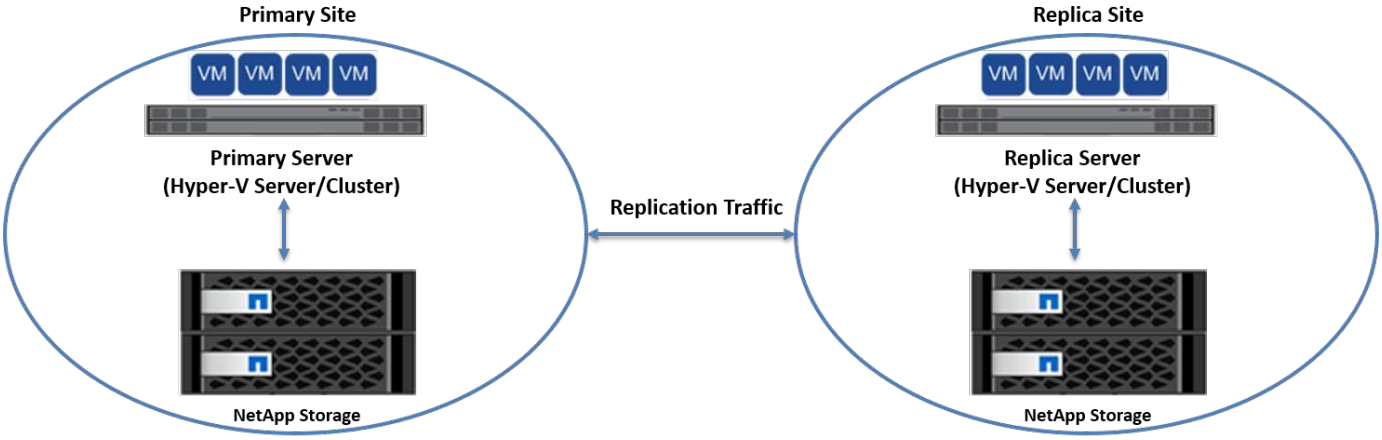


Lecture ultérieure

Pour plus d'informations sur le déploiement de la migration dynamique du stockage, reportez-vous à la section "Annexe E : déploiement de la migration dynamique du stockage Hyper-V.".

Hyper-V Replica : reprise après incident pour les machines virtuelles

Le réplica Hyper-V réplique les machines virtuelles Hyper-V depuis un site principal vers des machines virtuelles de réplica sur un site secondaire, assurant ainsi une reprise après incident asynchrone pour les machines virtuelles. Le serveur Hyper-V sur le site principal hébergeant les machines virtuelles est appelé serveur principal ; le serveur Hyper-V sur le site secondaire qui reçoit les machines virtuelles répliquées est appelé serveur de réplica. Un exemple de scénario de réplica Hyper-V est illustré dans la figure suivante. Vous pouvez utiliser Hyper-V Replica pour les machines virtuelles entre des serveurs Hyper-V faisant partie d'un cluster de basculement ou entre des serveurs Hyper-V indépendants qui ne font pas partie d'un cluster.



La réplication

Lorsque Hyper-V Replica est activé pour une machine virtuelle sur le serveur principal, la réplication initiale crée une machine virtuelle identique sur le serveur de réplica. Après la réplication initiale, Hyper-V Replica conserve un fichier journal pour les VHD de la machine virtuelle. Le fichier journal est relu dans l'ordre inverse vers le VHD de réplica en fonction de la fréquence de réplication. Ce journal et l'utilisation de l'ordre inverse permettent de s'assurer que les dernières modifications sont stockées et répliquées de manière asynchrone. Si la réplication ne se produit pas conformément à la fréquence attendue, une alerte est émise.

Réplication étendue

Hyper-V Replica prend en charge la réplication étendue dans laquelle un serveur de réplica secondaire peut être configuré pour la reprise après incident. Un serveur de réplica secondaire peut être configuré pour que le serveur de réplica reçoive les modifications sur les machines virtuelles de réplica. Dans un scénario de réplication étendue, les modifications apportées aux machines virtuelles primaires du serveur principal sont répliquées sur le serveur de réplica. Les modifications sont ensuite répliquées sur le serveur de réplica étendu. Les machines virtuelles peuvent être défaillantes vers le serveur de réplica étendu uniquement lorsque les serveurs principal et de réplica sont en panne.

Basculement

Le basculement n'est pas automatique. Le processus doit être déclenché manuellement. Il existe trois types de basculement :

- **Test failover.** ce type est utilisé pour vérifier qu'une machine virtuelle de réplica peut démarrer avec succès sur le serveur de réplica et qu'elle est lancée sur la machine virtuelle de réplica. Ce processus crée

une machine virtuelle de test en double lors du basculement, sans affecter la réplication de production normale.

- **Basculement planifié.** ce type est utilisé pour basculer les machines virtuelles pendant les temps d'arrêt planifiés ou les interruptions prévues. Ce processus est lancé sur la machine virtuelle principale, qui doit être désactivée sur le serveur principal avant l'exécution d'un basculement planifié. Après le basculement de la machine, Hyper-V Replica démarre la machine virtuelle de réplica sur le serveur de réplica.
- **Basculement non planifié.** ce type est utilisé lorsque des pannes inattendues se produisent. Ce processus est lancé sur la machine virtuelle de réplica et ne doit être utilisé que si la machine principale échoue.

Reprise après incident

Lorsque vous configurez la réplication pour une machine virtuelle, vous pouvez spécifier le nombre de points de restauration. Les points de restauration représentent des points dans le temps à partir desquels les données peuvent être récupérées à partir d'une machine répliquée.

Lecture ultérieure

- Pour plus d'informations sur le déploiement d'un réplica Hyper-V en dehors d'un environnement en cluster, reportez-vous à la section «"Déploiement d'un réplica Hyper-V en dehors d'un environnement en cluster"."
- Pour plus d'informations sur le déploiement d'un réplica Hyper-V dans un environnement en cluster, reportez-vous à la section «"Déployez le réplica Hyper-V dans un environnement en cluster"."

Efficacité du stockage

ONTAP offre une efficacité du stockage de pointe pour les environnements virtualisés tels que Microsoft Hyper-V. NetApp propose également des programmes de garantie d'efficacité du stockage.

Déduplication NetApp

La déduplication NetApp supprime les blocs dupliqués au niveau du volume de stockage et ne stocke qu'une seule copie physique, quel que soit le nombre de copies logiques présentes. Par conséquent, la déduplication crée l'illusion qu'il y a de nombreuses copies de ce bloc. La déduplication supprime automatiquement les blocs de données dupliqués au niveau d'un bloc de 4 Ko répartis sur un volume entier. Ce processus récupère le stockage pour réaliser des économies d'espace et de performances potentielles en réduisant le nombre d'écritures physiques sur le disque. La déduplication permet de réaliser des économies d'espace supérieures à 70 % dans les environnements Hyper-V.

Provisionnement fin

Le provisionnement fin constitue un moyen efficace de provisionner le stockage, car celui-ci n'est pas préalloué à l'avance. En d'autres termes, lorsqu'un volume ou une LUN est créé à l'aide du provisionnement fin, l'espace sur le système de stockage n'est pas utilisé. L'espace reste inutilisé jusqu'à ce que les données soient écrites sur la LUN ou le volume. Seul l'espace nécessaire pour stocker les données est utilisé. NetApp recommande d'activer le provisionnement fin sur le volume et de désactiver la réservation de LUN.

Qualité de service

La QoS du stockage dans NetApp ONTAP vous permet de regrouper des objets de stockage et de définir des limites de débit sur le groupe. La QoS du stockage peut être utilisée pour limiter le débit aux charges de travail et surveiller la performance des charges de travail. L'administrateur du stockage peut ainsi séparer les charges

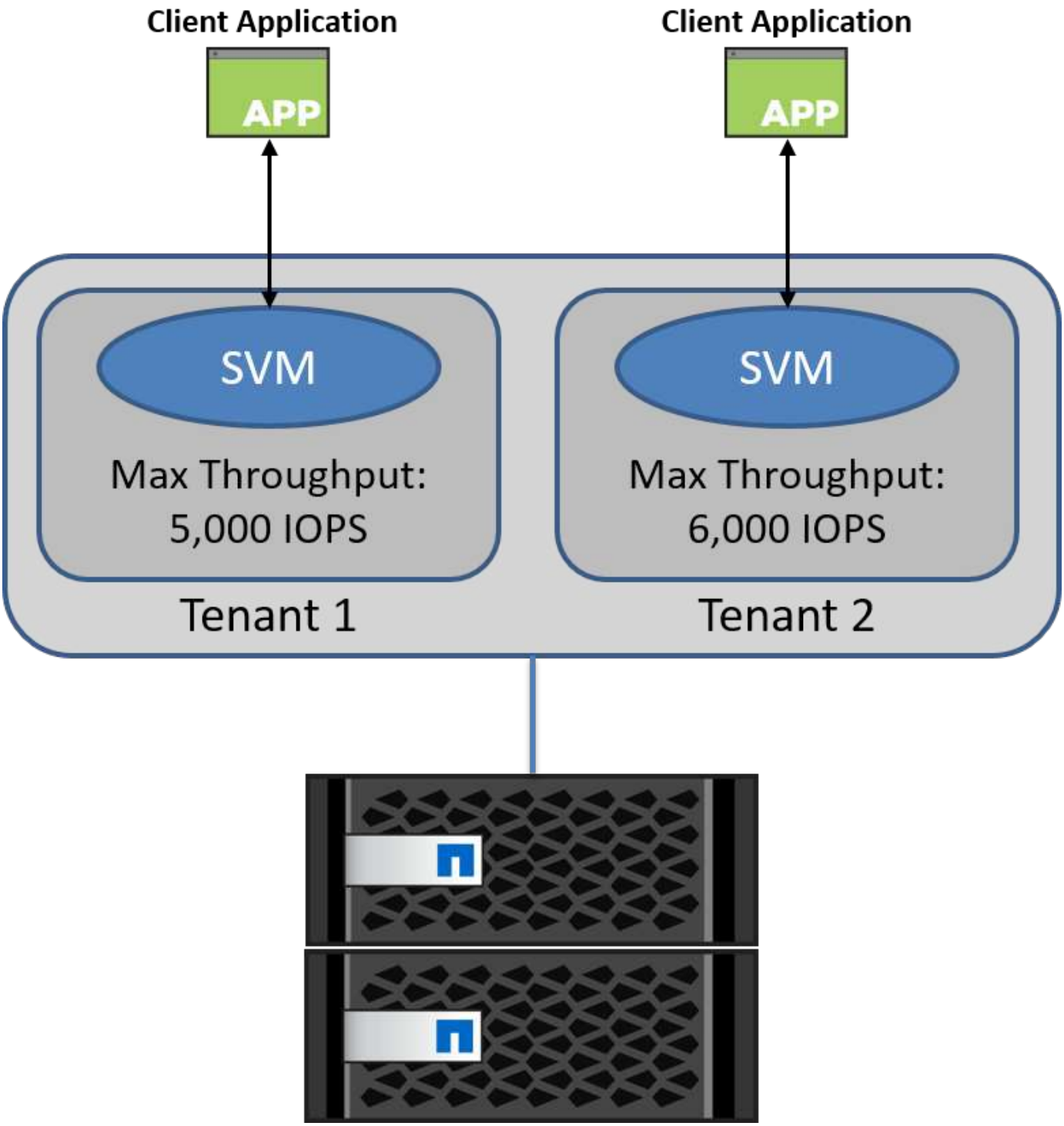
de travail par organisation, application, entité commerciale ou environnement de production ou de développement.

Dans les environnements d’entreprise, la QoS du stockage contribue à atteindre les objectifs suivants :

- Évitez que les charges de travail des utilisateurs ne s’entraffectent.
- Protège les applications stratégiques avec des temps de réponse spécifiques qui doivent être respectés dans les environnements IT à la demande (ITaaS).
- Empêche les locataires de s’entraffecter.
- Évite la dégradation des performances avec l’ajout de chaque nouveau locataire.

La QoS vous permet de limiter la quantité d’E/S envoyées à un SVM, un volume flexible, un LUN ou un fichier. Les E/S peuvent être limitées par le nombre d’opérations ou le débit brut.

La figure suivante illustre une SVM avec sa propre règle de QoS appliquée appliquant une limite de débit maximale.



Pour configurer un SVM avec ses propres politiques de QoS et surveiller le groupe de règles, exécutez les commandes suivantes sur votre cluster ONTAP :

```
# create a new policy group pg1 with a maximum throughput of 5,000 IOPS
cluster::> qos policy-group create pg1 -vserver vs1 -max-throughput
5000iops
```

```
# create a new policy group pg2 without a maximum throughput
cluster::> qos policy-group create pg2 -vserver vs2
```

```
# monitor policy group performance
cluster::> qos statistics performance show
```

```
# monitor workload performance
cluster::> qos statistics workload performance show
```

Sécurité

ONTAP fournit un système de stockage sécurisé pour le système d'exploitation Windows.

Antivirus Windows Defender

Windows Defender est un logiciel antimalware installé et activé par défaut sur Windows Server. Ce logiciel protège activement Windows Server contre les programmes malveillants connus et peut régulièrement mettre à jour les définitions d'antimalware via Windows Update. Les LUN NetApp et les partages SMB peuvent être analysés à l'aide de Windows Defender.

Lecture ultérieure

Pour plus d'informations, reportez-vous au "Présentation de Windows Defender".

BitLocker

Le chiffrement de lecteur BitLocker est une fonction de protection des données, suite à Windows Server 2012. Ce chiffrement protège les disques physiques, les LUN et les CSV.

Et des meilleures pratiques

Avant d'activer BitLocker, le CSV doit être mis en mode de maintenance. Par conséquent, NetApp recommande de prendre des décisions relatives à la sécurité basée sur BitLocker avant de créer des machines virtuelles sur le CSV afin d'éviter les temps d'arrêt.

Déploiement du serveur Nano

Découvrez comment déployer Microsoft Windows Nano Server.

Déploiement

Pour déployer un Nano Server en tant qu'hôte Hyper-V, procédez comme suit :

1. Connectez-vous à Windows Server en tant que membre du groupe d'administrateurs.
2. Copiez le dossier NanoServerImageGenerator du dossier \NanoServer dans l'ISO Windows Server sur le disque dur local.

3. Pour créer un Nano Server VHD/VHDX, procédez comme suit :
   a. Démarrez Windows PowerShell en tant qu'administrateur, accédez au dossier NanoServerImageGenerator copié sur le disque dur local et exécutez l'applet de commande suivante :

```
Set-ExecutionPolicy RemoteSigned
Import-Module .\NanoServerImageGenerator -Verbose
```

   b. Créez un VHD pour le Nano Server en tant qu'hôte Hyper-V en exécutant l'applet de commande PowerShell suivante. Cette commande vous invite à entrer un mot de passe administrateur pour le nouveau VHD.

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath <"input the path to the root of the contents of Windows
Server 2016 ISO"> -TargetPath <"input the path, including the
filename and extension where the resulting VHD/VHDX will be created">
-ComputerName <"input the name of the nano server computer you are
about to create"> -Compute
.. Dans l'exemple suivant, nous créons un disque dur virtuel Nano
Server avec la fonctionnalité hôte Hyper-V avec mise en cluster de
basculement activée. Cet exemple crée un disque dur virtuel Nano
Server à partir d'un fichier ISO monté à f:\. Le VHD nouvellement
créé est placé dans un dossier nommé NanoServer dans le dossier à
partir duquel l'applet de commande est exécutée. Le nom de
l'ordinateur est NanoServer et le VHD obtenu contient l'édition
standard de Windows Server.
```

```
 New-NanoServerImage -Edition Standard -DeploymentType Guest
-MediaPath f:\ -TargetPath .\NanoServer.vhd -ComputerName NanoServer
-Compute -Clustering
.. Avec l'applet de commande New-NanoServerImage, configurez les
paramètres qui définissent l'adresse IP, le masque de sous-réseau, la
passerelle par défaut, le serveur DNS, le nom de domaine, et ainsi de
suite.
```

4. Utilisez le VHD sur une machine virtuelle ou un hôte physique pour déployer Nano Server en tant qu'hôte Hyper-V :
   a. Pour le déploiement sur une machine virtuelle, créez une nouvelle machine virtuelle dans Hyper-V Manager et utilisez le VHD créé à l'étape 3.
   b. Pour le déploiement sur un hôte physique, copiez le VHD sur l'ordinateur physique et configurez-le pour qu'il démarre à partir de ce nouveau VHD. Tout d'abord, montez le VHD, exécutez bcdboot e:\Windows (où le VHD est monté sous E:\), démontez le VHD, redémarrez l'ordinateur physique et démarrez le Nano Server.
5. Connectez le Nano Server à un domaine (facultatif) :

a. Connectez-vous à n'importe quel ordinateur du domaine et créez un blob de données en exécutant l'applet de commande PowerShell suivante :

```
$domain = "<input the domain to which the Nano Server is to be
joined>"
$nanoserver = "<input name of the Nano Server>"
```

```
 djoin.exe /provision /domain $domain /machine $nanoserver /savefile
C:\temp\odjblob /reuse
.. Copiez le fichier odjblob sur le Nano Server en exécutant les
applets de commande PowerShell suivantes sur un ordinateur distant :
```

```
$nanoserver = "<input name of the Nano Server>"
$nanouname = ""<input username of the Nano Server>"
$nanopwd = ""<input password of the Nano Server>"
```

```
$filePath = 'c:\temp\odjblob'
$fileContents = Get-Content -Path $filePath -Encoding Unicode
```

```
$securenanopwd = ConvertTo-SecureString -AsPlainText -Force $nanopwd
$nanosecurecred = new-object management.automation.pscredential
$nanouname, $securenanopwd
```

```
Invoke-Command -VMName $nanoserver -Credential $nanosecurecred
-ArgumentList @($filePath,$fileContents) -ScriptBlock \{
    param($filePath,$data)
    New-Item -ItemType directory -Path c:\temp
    Set-Content -Path $filePath -Value $data -Encoding Unicode
    cd C:\temp
    djoin /requestodj /loadfile c:\temp\odjblob /windowspath
c:\windows /localos
}
```

b. Redémarrez le serveur Nano.

Connectez-vous au Nano Server

Pour vous connecter au Nano Server à distance à l'aide de PowerShell, procédez comme suit :

1. Ajoutez le Nano Server en tant qu'hôte de confiance sur l'ordinateur distant en exécutant l'applet de

commande suivante sur le serveur distant :

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts "<input IP Address of the
Nano Server>"
. Si l'environnement est sûr et si vous souhaitez définir tous les hôtes
à ajouter en tant qu'hôtes de confiance sur un serveur, exécutez la
commande suivante :
```

```
 Set-Item WSMan:\LocalHost\Client\TrustedHosts *
. Démarrez la session distante en exécutant l'applet de commande
suivante sur le serveur distant. Saisissez le mot de passe du Nano
Server lorsque vous y êtes invité.
```

```
Enter-PSSession -ComputerName "<input IP Address of the Nano Server>"
-Credential ~\Administrator
```

Pour vous connecter au Nano Server à distance à l'aide des outils de gestion de l'interface utilisateur graphique à partir d'un serveur Windows distant, exécutez les commandes suivantes :

1. Connectez-vous au serveur Windows en tant que membre du groupe d'administrateurs.
2. Démarrez Server Manager**.**
3. Pour gérer un Nano Server à distance à partir de Server Manager, cliquez avec le bouton droit de la souris sur tous les serveurs, cliquez sur Ajouter des serveurs, indiquez les informations du Nano Server et ajoutez-le. Vous pouvez maintenant voir le Nano Server dans la liste des serveurs. Sélectionnez le Nano Server, cliquez dessus avec le bouton droit de la souris et commencez à le gérer à l'aide des différentes options fournies.
4. Pour gérer les services sur un Nano Server à distance, procédez comme suit :
   a. Ouvrez Services dans la section Outils de Server Manager.
   b. Cliquez avec le bouton droit de la souris sur Services (local).
   c. Cliquez sur se connecter au serveur.
   d. Fournissez les détails du Nano Server pour afficher et gérer les services sur le Nano Server.
5. Si le rôle Hyper-V est activé sur le Nano Server, procédez comme suit pour le gérer à distance à partir d'Hyper-V Manager :
   a. Ouvrez Hyper-V Manager dans la section Outils de Server Manager.
   b. Cliquez avec le bouton droit de la souris sur Gestionnaire Hyper-V.
   c. Cliquez sur se connecter au serveur et indiquez les détails du Nano Server. Le Nano Server peut désormais être géré en tant que serveur Hyper-V pour créer et gérer des machines virtuelles.
6. Si le rôle de mise en cluster de basculement est activé sur le Nano Server, procédez comme suit pour le gérer à distance à partir du gestionnaire de cluster de basculement :
   a. Ouvrez le Gestionnaire de clusters de basculement à partir de la section Outils de Server Manager.
   b. Effectuez des opérations de mise en cluster avec le Nano Server.

Déployez le cluster Hyper-V.

Cette annexe décrit le déploiement d'un cluster Hyper-V.

Prérequis

- Au moins deux serveurs Hyper-V sont connectés l'un à l'autre.
- Au moins un commutateur virtuel est configuré sur chaque serveur Hyper-V.
- La fonctionnalité cluster de basculement est activée sur chaque serveur Hyper-V.
- Les partages SMB ou les CSV sont utilisés comme stockage partagé pour stocker les machines virtuelles et leurs disques pour la mise en cluster Hyper-V.
- Le stockage ne doit pas être partagé entre des clusters différents. Vous ne devez avoir qu'un seul partage CSV/CIFS par cluster.
- Si le partage SMB est utilisé comme stockage partagé, les autorisations sur le partage SMB doivent être configurées de manière à accorder l'accès aux comptes ordinateur de tous les serveurs Hyper-V du cluster.

Déploiement

1. Connectez-vous à l'un des serveurs Windows Hyper-V en tant que membre du groupe d'administrateurs.
2. Démarrez Server Manager**.**
3. Dans la section Outils, cliquez sur Gestionnaire de clusters de basculement.
4. Cliquez sur le menu Créer un cluster à partir des actions.
5. Fournir des détails sur le serveur Hyper-V qui fait partie de ce cluster.
6. Validez la configuration du cluster. Sélectionnez Oui lorsque vous êtes invité à valider la configuration du cluster, puis sélectionnez les tests requis pour vérifier si les serveurs Hyper-V satisfont aux conditions préalables requises pour faire partie du cluster.
7. Une fois la validation réussie, l'assistant de création de cluster démarre. Dans l'assistant, indiquez le nom du cluster et l'adresse IP du nouveau cluster. Un nouveau cluster de basculement est ensuite créé pour le serveur Hyper-V.
8. Cliquez sur le nouveau cluster créé dans Failover Cluster Manager et gérez-le.
9. Définir le stockage partagé à utiliser par le cluster. Il peut s'agir d'un partage SMB ou d'un fichier CSV.
10. L'utilisation d'un partage SMB comme stockage partagé ne nécessite pas d'étapes spéciales.
    - Configurez un partage CIFS sur un contrôleur de stockage NetApp. Pour ce faire, reportez-vous à la section «"Le provisionnement dans les environnements SMB"«.
11. Pour utiliser un fichier CSV comme stockage partagé, procédez comme suit :
    a. Configurer les LUN sur un contrôleur de stockage NetApp Pour ce faire, reportez-vous à la section « provisionnement dans les environnements SAN ».
    b. Assurez-vous que tous les serveurs Hyper-V du cluster de basculement peuvent voir les LUN NetApp. Pour ce faire pour tous les serveurs Hyper-V faisant partie du cluster de basculement, assurez-vous que leurs initiateurs sont ajoutés au groupe initiateur sur le stockage NetApp. Assurez-vous également que leurs LUN sont détectées et que MPIO est activé.
    c. Sur l'un des serveurs Hyper-V du cluster, effectuez les opérations suivantes :
        i. Mettre la LUN en ligne, initialiser le disque, créer un nouveau volume simple et le formater à l'aide de NTFS ou de ReFS.

ii. Dans le Gestionnaire de clusters de basculement, développez le cluster, développez stockage, cliquez avec le bouton droit de la souris sur disques, puis cliquez sur Ajouter des disques. L'assistant Ajouter des disques à un cluster s'ouvre alors et affiche la LUN comme disque. Cliquez sur OK pour ajouter la LUN en tant que disque.

iii. La LUN s'appelle désormais « disque en cluster » et est affichée comme stockage disponible sous disques.

d. Cliquez avec le bouton droit de la souris sur la LUN (disque en cluster), puis cliquez sur Ajouter aux volumes partagés du cluster. La LUN s'affiche désormais au format CSV.

e. Le CSV est visible et accessible simultanément depuis tous les serveurs Hyper-V du cluster de basculement à son emplacement local C:\ClusterStorage\.

12. Créer un serveur virtuel hautement disponible :

    a. Dans Failover Cluster Manager, sélectionnez et développez le cluster créé précédemment.

    b. Cliquez sur rôles, puis sur machines virtuelles dans actions. Cliquez sur Nouvelle machine virtuelle.

    c. Sélectionnez dans le cluster le nœud sur lequel la machine virtuelle doit résider.

    d. Dans l'assistant Virtual machine Creation (création d'une machine virtuelle), indiquez le stockage partagé (partage SMB ou CSV) comme chemin d'accès pour stocker la machine virtuelle et ses disques.

    e. Utilisez Hyper-V Manager pour définir le stockage partagé (partage SMB ou CSV) comme chemin par défaut pour stocker la machine virtuelle et ses disques pour un serveur Hyper-V.

13. Test du basculement planifié. Déplacez des machines virtuelles vers un autre nœud via une migration dynamique, une migration rapide ou une migration du stockage (déplacement). Révision "Migration dynamique dans un environnement en cluster" pour en savoir plus.

14. Tester le basculement non planifié. Arrêtez le service de cluster sur le serveur propriétaire de la machine virtuelle.

Déployer Hyper-V Live migration dans un environnement en cluster

Cette annexe décrit le déploiement de la migration dynamique dans un environnement en cluster.

Prérequis

Pour déployer la migration en direct, vous devez configurer des serveurs Hyper-V dans un cluster de basculement avec stockage partagé. Révision "Déployez le cluster Hyper-V." pour en savoir plus.

Déploiement

Pour utiliser la migration dynamique dans un environnement en cluster, procédez comme suit :

1. Dans Failover Cluster Manager, sélectionnez et développez le cluster. Si le cluster n'est pas visible, cliquez sur Gestionnaire de cluster de basculement, cliquez sur se connecter au cluster et indiquez le nom du cluster.
2. Cliquez sur rôles, qui répertorie toutes les machines virtuelles disponibles dans un cluster.
3. Cliquez avec le bouton droit de la souris sur la machine virtuelle et cliquez sur déplacer. Vous disposez ainsi de trois options :
    - **Migration dynamique.** vous pouvez sélectionner un nœud manuellement ou autoriser le cluster à sélectionner le meilleur nœud. Dans le cadre de la migration dynamique, le cluster copie la mémoire

utilisée par la machine virtuelle du nœud actuel vers un autre nœud. Par conséquent, lorsque la machine virtuelle est migrée vers un autre nœud, la mémoire et les informations d'état requises par la machine virtuelle sont déjà en place pour cette dernière. Cette méthode de migration est quasi instantanée, mais une seule machine virtuelle peut être migrée en direct à la fois.

- **Migration rapide.** vous pouvez sélectionner un nœud manuellement ou autoriser le cluster à sélectionner le meilleur nœud. Lors d'une migration rapide, le cluster copie la mémoire utilisée par une machine virtuelle sur un disque du système de stockage. Par conséquent, lorsque la machine virtuelle est migrée vers un autre nœud, l'autre nœud peut rapidement lire la mémoire et les informations d'état requises par la machine virtuelle à partir du disque. Grâce à une migration rapide, plusieurs machines virtuelles peuvent être migrées simultanément.
- **Migration du stockage de la machine virtuelle.** cette méthode utilise l'assistant de déplacement du stockage de la machine virtuelle. Cet assistant vous permet de sélectionner le disque de la machine virtuelle ainsi que d'autres fichiers à déplacer vers un autre emplacement, qui peut être un partage CSV ou SMB.

Déployer Hyper-V Live migration en dehors d'un environnement en cluster

Cette section décrit le déploiement de la migration dynamique Hyper-V en dehors d'un environnement en cluster.

Prérequis

- Serveurs Hyper-V autonomes avec stockage indépendant ou stockage SMB partagé.
- Rôle Hyper-V installé à la fois sur les serveurs source et de destination.
- Les deux serveurs Hyper-V appartiennent au même domaine ou aux domaines qui se font confiance.

Déploiement

Pour effectuer une migration en direct dans un environnement non mis en cluster, configurez les serveurs Hyper-V source et de destination afin qu'ils puissent envoyer et recevoir des opérations de migration en direct. Sur les deux serveurs Hyper-V, procédez comme suit :

1. Ouvrez Hyper-V Manager dans la section Outils de Server Manager.
2. Dans actions, cliquez sur Paramètres Hyper-V.
3. Cliquez sur migrations dynamiques et sélectionnez Activer les migrations dynamiques entrantes et sortantes.
4. Choisissez d'autoriser le trafic de migration en direct sur n'importe quel réseau disponible ou uniquement sur des réseaux spécifiques.
5. Vous pouvez également configurer le protocole d'authentification et les options de performances à partir de la section Avancé de Live migrations.
6. Si CredSSP est utilisé comme protocole d'authentification, assurez-vous de vous connecter au serveur Hyper-V source à partir du serveur Hyper-V de destination avant de déplacer la machine virtuelle.
7. Si Kerberos est utilisé comme protocole d'authentification, configurez la délégation contrainte. Pour ce faire, vous devez accéder au contrôleur de domaine Active Directory. Pour configurer la délégation, procédez comme suit :
   a. Connectez-vous au contrôleur de domaine Active Directory en tant qu'administrateur.
   b. Démarrez Server Manager.
   c. Dans la section Outils, cliquez sur utilisateurs et ordinateurs Active Directory.

d. Développez le domaine et cliquez sur ordinateurs.
e. Sélectionnez le serveur Hyper-V source dans la liste, cliquez dessus avec le bouton droit de la souris et cliquez sur Propriétés.
f. Dans l'onglet délégation, sélectionnez faire confiance à cet ordinateur pour la délégation aux services spécifiés uniquement.
g. Sélectionnez utiliser Kerberos uniquement.
h. Cliquez sur Ajouter pour ouvrir l'assistant Ajouter des services.
i. Dans Ajouter des services, cliquez sur utilisateurs et ordinateurs, ce qui ouvre Sélectionner utilisateurs ou ordinateurs**.**
j. Indiquez le nom du serveur Hyper-V de destination et cliquez sur OK.
   - Pour déplacer le stockage de la machine virtuelle, sélectionnez CIFS.
   - Pour déplacer des machines virtuelles, sélectionnez le service Microsoft Virtual System migration.
k. Dans l'onglet délégation, cliquez sur OK.
l. Dans le dossier ordinateurs, sélectionnez le serveur Hyper-V de destination dans la liste et répétez le processus. Dans Sélectionner utilisateurs ou ordinateurs, indiquez le nom du serveur Hyper-V source.

8. Déplacer la VM.
   a. Ouvrez Hyper-V Manager.
   b. Cliquez avec le bouton droit de la souris sur une machine virtuelle et cliquez sur déplacer.
   c. Choisissez déplacer la machine virtuelle.
   d. Spécifier le serveur Hyper-V de destination pour la machine virtuelle.
   e. Choisissez les options de déplacement. Pour Shared Live migration, choisissez déplacer uniquement la machine virtuelle. Pour Shared Nothing Live migration, choisissez l'une des deux autres options en fonction de vos préférences.
   f. Indiquez l'emplacement de la machine virtuelle sur le serveur Hyper-V de destination en fonction de vos préférences.
   g. Vérifiez le récapitulatif et cliquez sur OK pour déplacer la machine virtuelle.

Déployez Hyper-V Storage Live migration

Découvrez comment configurer la migration dynamique du stockage Hyper-V.

Prérequis

- Vous devez disposer d'un serveur Hyper-V autonome avec stockage indépendant (DAS ou LUN) ou d'un stockage SMB (local ou partagé entre d'autres serveurs Hyper-V).
- Le serveur Hyper-V doit être configuré pour la migration en direct. Passez en revue la section sur le déploiement dans "Migration dynamique en dehors d'un environnement en cluster".

Déploiement

1. Ouvrez Hyper-V Manager.
2. Cliquez avec le bouton droit de la souris sur une machine virtuelle et cliquez sur déplacer.
3. Sélectionnez déplacer le stockage de l'ordinateur virtuel.

4. Sélectionnez les options de déplacement du stockage en fonction de vos préférences.
5. Indiquez le nouvel emplacement des éléments de la machine virtuelle.
6. Vérifiez le récapitulatif et cliquez sur OK pour déplacer le stockage de la machine virtuelle.

Déployer le réplica Hyper-V en dehors d'un environnement en cluster

Cette annexe décrit le déploiement d'un réplica Hyper-V en dehors d'un environnement en cluster.

Prérequis

- Vous avez besoin de serveurs Hyper-V autonomes situés dans le même emplacement géographique ou dans des emplacements distincts servant de serveurs principaux et de serveurs de réplica.
- Si des sites distincts sont utilisés, le pare-feu de chaque site doit être configuré pour permettre la communication entre les serveurs principal et de réplica.
- Le serveur de réplica doit disposer d'un espace suffisant pour stocker les charges de travail répliquées.

Déploiement

1. Configurez le serveur de réplica.
   a. Pour que les règles de pare-feu entrantes autorisent le trafic de réplication entrant, exécutez l'applet de commande PowerShell suivante :

```
 Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP Listener
(TCP-In)"
.. Ouvrez Hyper-V Manager dans la section Outils de Server Manager.
.. Cliquez sur Paramètres Hyper-V dans actions.
.. Cliquez sur Configuration de la réplication et sélectionnez
Activer cet ordinateur en tant que serveur de réplica.
.. Dans la section authentification et ports, sélectionnez la méthode
et le port d'authentification.
.. Dans la section autorisation et stockage, spécifiez l'emplacement
où stocker les machines virtuelles et les fichiers répliqués.
```

2. Activez la réplication des machines virtuelles sur le serveur principal. La réplication des machines virtuelles est activée par machine virtuelle, et non pour l'ensemble du serveur Hyper-V.
   a. Dans Hyper-V Manager, cliquez avec le bouton droit de la souris sur un serveur virtuel, puis cliquez sur Activer la réplication pour ouvrir l'assistant Activer la réplication.
   b. Indiquez le nom du serveur de réplica sur lequel la machine virtuelle doit être répliquée.
   c. Indiquez le type d'authentification et le port du serveur de réplica qui a été configuré pour recevoir le trafic de réplication sur le serveur de réplica.
   d. Sélectionnez les VHD à répliquer.
   e. Choisissez la fréquence (durée) à laquelle les modifications sont envoyées au serveur de réplica.
   f. Configurez les points de récupération pour spécifier le nombre de points de récupération à conserver sur le serveur de réplica.

g. Choisissez méthode de réplication initiale pour spécifier la méthode de transfert de la copie initiale des données de la machine virtuelle vers le serveur de réplica.

h. Vérifiez le résumé et cliquez sur Terminer.

i. Ce processus crée une réplique de machine virtuelle sur le serveur de réplica.

La réplication

1. Exécutez un basculement de test pour vous assurer que la machine virtuelle de réplica fonctionne correctement sur le serveur de réplica. Le test crée une machine virtuelle temporaire sur le serveur de réplica.
   a. Connectez-vous au serveur de réplica.
   b. Dans Hyper-V Manager, cliquez avec le bouton droit de la souris sur un serveur virtuel de réplica, cliquez sur réplication, puis sur Test Failover.
   c. Choisissez le point de restauration à utiliser.
   d. Ce processus crée un VM du même nom ajouté à -Test.
   e. Vérifier la machine virtuelle pour s'assurer que tout fonctionne correctement.
   f. Après le basculement, la machine virtuelle de test de réplica est supprimée si vous sélectionnez Arrêter le basculement de test pour elle.
2. Exécutez un basculement planifié pour répliquer les dernières modifications sur la machine virtuelle principale vers la machine virtuelle de réplica.
   a. Connectez-vous au serveur principal.
   b. Désactivez la machine virtuelle à basculer.
   c. Dans Hyper-V Manager, cliquez avec le bouton droit de la souris sur le serveur virtuel désactivé, cliquez sur réplication, puis sur basculement planifié.
   d. Cliquez sur basculement pour transférer les dernières modifications apportées à la machine virtuelle vers le serveur de réplica.
3. Exécuter un basculement non planifié en cas de défaillance de la machine virtuelle principale.
   a. Connectez-vous au serveur de réplica.
   b. Dans Hyper-V Manager, cliquez avec le bouton droit de la souris sur un serveur virtuel de réplica, cliquez sur réplication, puis sur basculement.
   c. Choisissez le point de restauration à utiliser.
   d. Cliquez sur basculement pour basculer le serveur virtuel.

Déployer la réplique Hyper-V dans un environnement en cluster

Découvrez comment déployer et configurer une réplique Hyper-V avec le cluster de basculement Windows Server.

Prérequis

- Vous devez disposer de clusters Hyper-V situés dans le même emplacement ou dans des emplacements géographiques distincts, et servant de clusters principal et de clusters de réplica. Révision "Déployez le cluster Hyper-V." pour en savoir plus.
- Si des sites distincts sont utilisés, le pare-feu de chaque site doit être configuré pour permettre la communication entre les clusters principal et de réplica.

- Le cluster de réplica doit disposer d'un espace suffisant pour stocker les charges de travail répliquées.

Déploiement

1. Activez les règles de pare-feu sur tous les nœuds d'un cluster. Exécutez l'applet de commande PowerShell suivante avec des privilèges d'administrateur sur tous les nœuds des clusters principal et de réplica.

```
# For Kerberos authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica HTTP
Listener (TCP-In)"}}
```

```
# For Certificate authentication
get-clusternode | ForEach-Object \{Invoke-command -computername $_.name
-scriptblock \{Enable-Netfirewallrule -displayname "Hyper-V Replica
HTTPS Listener (TCP-In)"}}
```

2. Configurer le cluster de réplica.
   a. Configurez le courtier de réplica Hyper-V avec un nom NetBIOS et une adresse IP à utiliser comme point de connexion au cluster utilisé comme cluster de réplica.
      i. Ouvrez Failover Cluster Manager.
      ii. Développez le cluster, cliquez sur rôles, puis sur le volet configurer le rôle à partir des actions.
      iii. Sélectionnez Hyper-V Replica Broker dans la page Sélectionner un rôle.
      iv. Indiquez le nom NetBIOS et l'adresse IP à utiliser comme point de connexion au cluster (point d'accès client).
      v. Ce processus crée un rôle de courtier de réplica Hyper-V. Vérifiez qu'elle est bien en ligne.
   b. Configurer les paramètres de réplication.
      i. Cliquez avec le bouton droit de la souris sur le courtier de répliques créé lors des étapes précédentes, puis cliquez sur Paramètres de réplication.
      ii. Sélectionnez Activer ce cluster en tant que serveur de réplica.
      iii. Dans la section authentification et ports, sélectionnez la méthode et le port d'authentification.
      iv. Dans la section autorisation et stockage, sélectionnez les serveurs autorisés à répliquer des machines virtuelles sur ce cluster. Spécifiez également l'emplacement par défaut où les VM répliquées sont stockées.

La réplication

La réplication est similaire au processus décrit dans la section "Réplique hors d'un environnement en cluster".

Où trouver des informations complémentaires

Ressources supplémentaires pour Microsoft Windows et Hyper-V.

- Concepts relatifs à ONTAP

https://docs.netapp.com/us-en/ontap/concepts/introducing-data-management-software-concept.html

- Bonnes pratiques pour le SAN moderne
  https://www.netapp.com/media/10680-tr4080.pdf
- Disponibilité et intégrité des données des baies SAN 100 % Flash de NetApp avec NetApp ASA
  https://www.netapp.com/pdf.html?item=/media/85671-tr-4968.pdf
- Documentation SMB
  https://docs.netapp.com/us-en/ontap/smb-admin/index.html
- Mise en route avec Nano Server
  https://technet.microsoft.com/library/mt126167.aspx
- Nouveautés d'Hyper-V sur Windows Server
  https://technet.microsoft.com/windows-server-docs/compute/hyper-v/what-s-new-in-hyper-v-on-windows

Microsoft SQL Server

Présentation

ONTAP propose une solution de sécurité et de performances pour vos bases de données Microsoft SQL Server, tout en fournissant des outils de pointe pour gérer votre environnement.

> Cette documentation remplace le rapport technique *TR-4590 : guide des meilleures pratiques pour Microsoft SQL Server avec ONTAP*

NetApp suppose que le lecteur a une connaissance pratique des éléments suivants :

- ONTAP
- Architecture et administration de Microsoft SQL Server
- NetApp SnapCenter en tant que logiciel de sauvegarde, qui inclut :
  - Plug-in SnapCenter pour Microsoft Windows
  - Plug-in SnapCenter pour SQL Server

La portée de cette section sur les meilleures pratiques se limite à la conception technique basée sur les principes et les normes préconisant NetApp pour l'infrastructure de stockage. L'implémentation de bout en bout n'est pas concernée.

Pour plus d'informations sur la compatibilité entre les produits NetApp, consultez le "Matrice d'interopérabilité NetApp (IMT)".

Workloads Microsoft SQL Server

Avant de déployer SQL Server, vous devez comprendre les exigences de charge de travail des applications prises en charge par vos instances de base de données SQL Server. Chaque application a des exigences variables en termes de capacité, de performance et de disponibilité. Par conséquent, chaque base de données doit être conçue de manière à répondre de manière optimale à ces exigences. De nombreuses entreprises classent les bases de données en plusieurs niveaux de gestion, en utilisant les exigences des applications pour définir des contrats de niveau de service. Les charges de travail SQL Server sont souvent classées comme décrit ci-dessous :

- OLTP, qui sont souvent les bases de données les plus stratégiques d'une entreprise. Ces bases de données prennent généralement en charge les applications orientées client et sont considérées comme essentielles aux opérations stratégiques de l'entreprise. Les bases de données OLTP stratégiques et les applications qu'elles prennent en charge respectent souvent des SLA qui exigent des performances élevées, qui sont sensibles à la dégradation des performances et qui exigent une disponibilité maximale. Ils peuvent également être candidats pour toujours sur les clusters de basculement ou pour toujours sur les groupes de disponibilité. La combinaison d'E/S sur ces types de bases de données se caractérise généralement par une lecture aléatoire de 75 à 90 % et une écriture de 25 à 10 %.
- Bases de données du système d'aide à la décision (DSS), parfois appelées entrepôts de données. Ces bases de données jouent un rôle stratégique pour de nombreuses entreprises qui s'appuient sur l'analytique pour leurs activités. Ces bases de données sont sensibles à l'utilisation du CPU et aux opérations de lecture à partir du disque lors de l'exécution de requêtes. Dans de nombreuses entreprises, les bases de données DSS sont les plus critiques à la fin du mois, du trimestre et de l'année Ce workload

présente généralement un mélange d'E/S de lecture de près de 100 %, et le débit d'E/S est souvent plus important que les IOPS.

Configuration de la base de données

Configuration du processeur

Les performances de SQL Server dépendent de plusieurs unités centrales et de la configuration principale.

Hyper-Threading

L'Hyper-threading fait référence à la mise en œuvre simultanée de plusieurs threads (SMT), qui améliore la parallélisation des calculs réalisés sur les processeurs x86. La colocation sécurisée est disponible sur les processeurs Intel et AMD.

L'Hyper-threading entraîne des CPU logiques qui apparaissent sous forme de CPU physiques au système d'exploitation. SQL Server voit ensuite ces CPU supplémentaires et les utilise comme s'il y avait plus de cœurs que physiquement présents. Cela peut considérablement améliorer les performances en augmentant la parallélisation.

La mise en garde ici est que chaque version de SQL Server a ses propres limites sur la puissance de calcul qu'il peut utiliser. Pour plus d'informations, voir "Limites de capacité de calcul par édition de SQL Server".

Cœurs et gestion des licences

Il existe deux options de licence pour SQL Server. Le premier est connu sous le nom de modèle serveur + licence d'accès client (CAL) ; le second est le modèle par cœur de processeur. Bien que vous puissiez accéder à toutes les fonctionnalités du produit disponibles dans SQL Server avec la stratégie serveur + CAL, il existe une limite matérielle de 20 cœurs de processeur par socket. Même si vous disposez de SQL Server Enterprise Edition + CAL pour un serveur avec plus de 20 cœurs de processeur par socket, l'application ne peut pas utiliser tous ces cœurs à la fois sur cette instance.

L'image ci-dessous montre le message du journal SQL Server après le démarrage indiquant l'application de la limite de noyau.

```
2017-01-11 07:16:30.71 Server      Microsoft SQL Server 2016
(RTM) - 13.0.1601.5 (X64)
     Apr 29 2016 23:23:58
     Copyright (c) Microsoft Corporation
     Enterprise Edition (64-bit) on Windows Server 2016
Datacenter 6.3 <X64> (Build 14393: )

2017-01-11 07:16:30.71 Server      UTC adjustment: -8:00
2017-01-11 07:16:30.71 Server      (c) Microsoft Corporation.
2017-01-11 07:16:30.71 Server      All rights reserved.
2017-01-11 07:16:30.71 Server      Server process ID is 10176.
2017-01-11 07:16:30.71 Server      System Manufacturer:
'FUJITSU', System Model: 'PRIMERGY RX2540 M1'.
2017-01-11 07:16:30.71 Server      Authentication mode is MIXED.
2017-01-11 07:16:30.71 Server      Logging SQL Server messages
in file 'C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG'.
2017-01-11 07:16:30.71 Server      The service account is 'SEA-
TM\FUJIA2R30$'. This is an informational message; no user action
is required.
2017-01-11 07:16:30.71 Server      Registry startup parameters:
       -d C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\master.mdf
       -e C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\Log\ERRORLOG
       -l C:\Program Files\Microsoft SQL Server
\MSSQL13.MSSQLSERVER\MSSQL\DATA\mastlog.ldf
       -T 3502
       -T 834
2017-01-11 07:16:30.71 Server      Command Line Startup
Parameters:
       -s "MSSQLSERVER"
2017-01-11 07:16:30.72 Server      SQL Server detected 2 sockets
with 18 cores per socket and 36 logical processors per socket,
72 total logical processors; using 40 logical processors based
on SQL Server licensing. This is an informational message; no
user action is required.
2017-01-11 07:16:30.72 Server      SQL Server is starting at
```

Par conséquent, pour utiliser tous les CPU, vous devez utiliser la licence par cœur de processeur. Pour plus d'informations sur les licences SQL Server, reportez-vous à la section "SQL Server 2022 : une plateforme de données moderne".

Affinité CPU

Il est peu probable que vous ayez à modifier les valeurs par défaut de l'affinité du processeur à moins que vous ne rencontriez des problèmes de performances, mais il est toujours utile de comprendre ce qu'elles sont et comment elles fonctionnent.

SQL Server prend en charge l'affinité de processeur par deux options :

- Masque d'affinité du processeur
- Masque d'E/S d'affinité

SQL Server utilise tous les processeurs disponibles dans le système d'exploitation (si la licence par processeur est choisie). Il crée également des planificateurs f0r chaque CPU pour optimiser l'utilisation des ressources pour une charge de travail donnée. En mode multitâche, le système d'exploitation ou d'autres applications du serveur peuvent basculer les threads de traitement d'un processeur à un autre. SQL Server est une application qui consomme beaucoup de ressources et les performances peuvent en être affectées. Pour minimiser l'impact, vous pouvez configurer les processeurs de sorte que toute la charge SQL Server soit dirigée vers un groupe de processeurs présélectionné. Pour ce faire, utilisez le masque d'affinité du

processeur.

L'option de masque d'E/S d'affinité lie les E/S de disque SQL Server à un sous-ensemble de processeurs. Dans les environnements OLTP SQL Server, cette extension peut considérablement améliorer les performances des threads SQL Server exécutant des opérations d'E/S.

Degré maximal de parallélisme (MAXDOP)

Par défaut, SQL Server utilise tous les CPU disponibles pendant l'exécution d'une requête si la licence par cœur de processeur est choisie.

Bien que cela soit utile pour les requêtes volumineuses, il peut causer des problèmes de performances et limiter la simultanéité. Une meilleure approche consiste à limiter le parallélisme au nombre de cœurs physiques dans un seul socket de processeur. Par exemple, sur un serveur doté de deux sockets CPU physiques avec 12 cœurs par socket, quel que soit l'hyper-threading, `MAXDOP` doit être défini sur 12. `MAXDOP` Impossible de restreindre ou de dicter le CPU à utiliser. Elle limite le nombre de processeurs pouvant être utilisés par une seule requête de lot.

> **NetApp recommande** pour DSS comme les data warehouses, commencez par `MAXDOP` 50 et explorez le réglage vers le haut ou vers le bas si nécessaire. Assurez-vous de mesurer les requêtes critiques dans votre application lorsque vous effectuez des modifications.

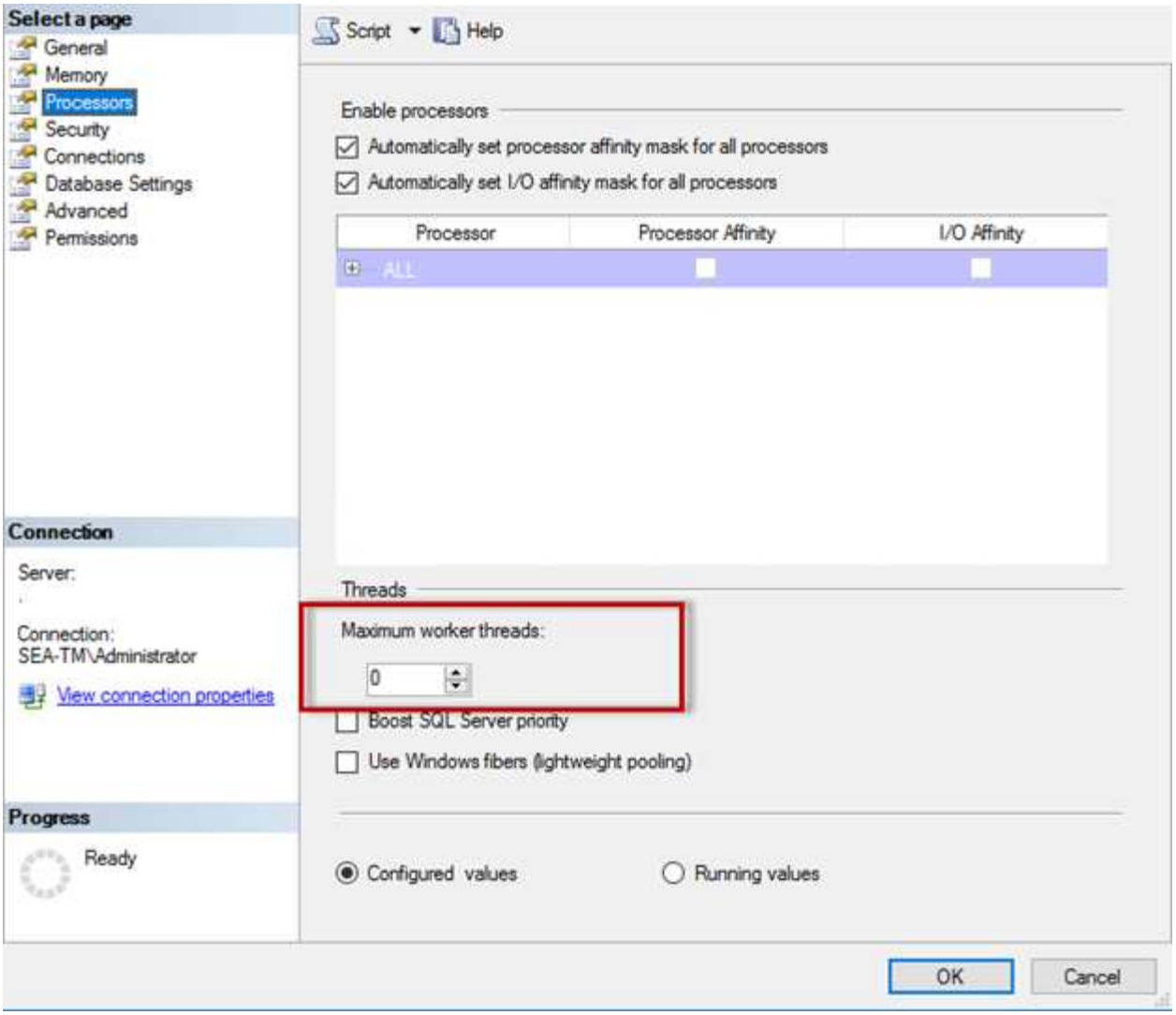
Nombre max. De threads de travail

L'option max worker threads permet d'optimiser les performances lorsqu'un grand nombre de clients sont connectés à SQL Server.

Normalement, un thread de système d'exploitation distinct est créé pour chaque requête. Si des centaines de connexions simultanées sont effectuées sur SQL Server, la configuration à un thread par requête peut consommer un excès de ressources système. Cette `max worker threads` option permet d'améliorer les performances en permettant à SQL Server de créer un pool de threads de travail pouvant traiter collectivement un plus grand nombre de requêtes.

La valeur par défaut est 0, ce qui permet à SQL Server de configurer automatiquement le nombre de threads de travail au démarrage. Cela fonctionne pour la plupart des systèmes. Max worker threads est une option avancée qui ne doit pas être modifiée sans l'aide d'un administrateur de base de données expérimenté (DBA).

Quand devez-vous configurer SQL Server pour utiliser davantage de threads de travail ? Si la longueur moyenne de la file d'attente de travail de chaque planificateur est supérieure à 1, vous pouvez bénéficier de l'ajout de threads supplémentaires au système, mais uniquement si la charge n'est pas liée au processeur ou si d'autres files d'attente importantes sont en cours. Si l'une ou l'autre de ces opérations se produit, l'ajout de threads n'est pas utile, car ils finissent par attendre les autres goulets d'étranglement du système. Pour plus d'informations sur le nombre maximal de threads de travail, reportez-vous à la section "Configurez l'option de configuration du serveur de threads de travail max".



Configuration du nombre maximal de threads de travail à l'aide de SQL Server Management Studio.

L'exemple suivant montre comment configurer l'option max Work threads à l'aide de T-SQL.

```
EXEC sp_configure 'show advanced options', 1;
GO
RECONFIGURE ;
GO
EXEC sp_configure 'max worker threads', 900 ;
GO
RECONFIGURE;
GO
```

Configuration de la mémoire

La section suivante décrit les paramètres de mémoire SQL Server requis pour optimiser les performances de la base de données.

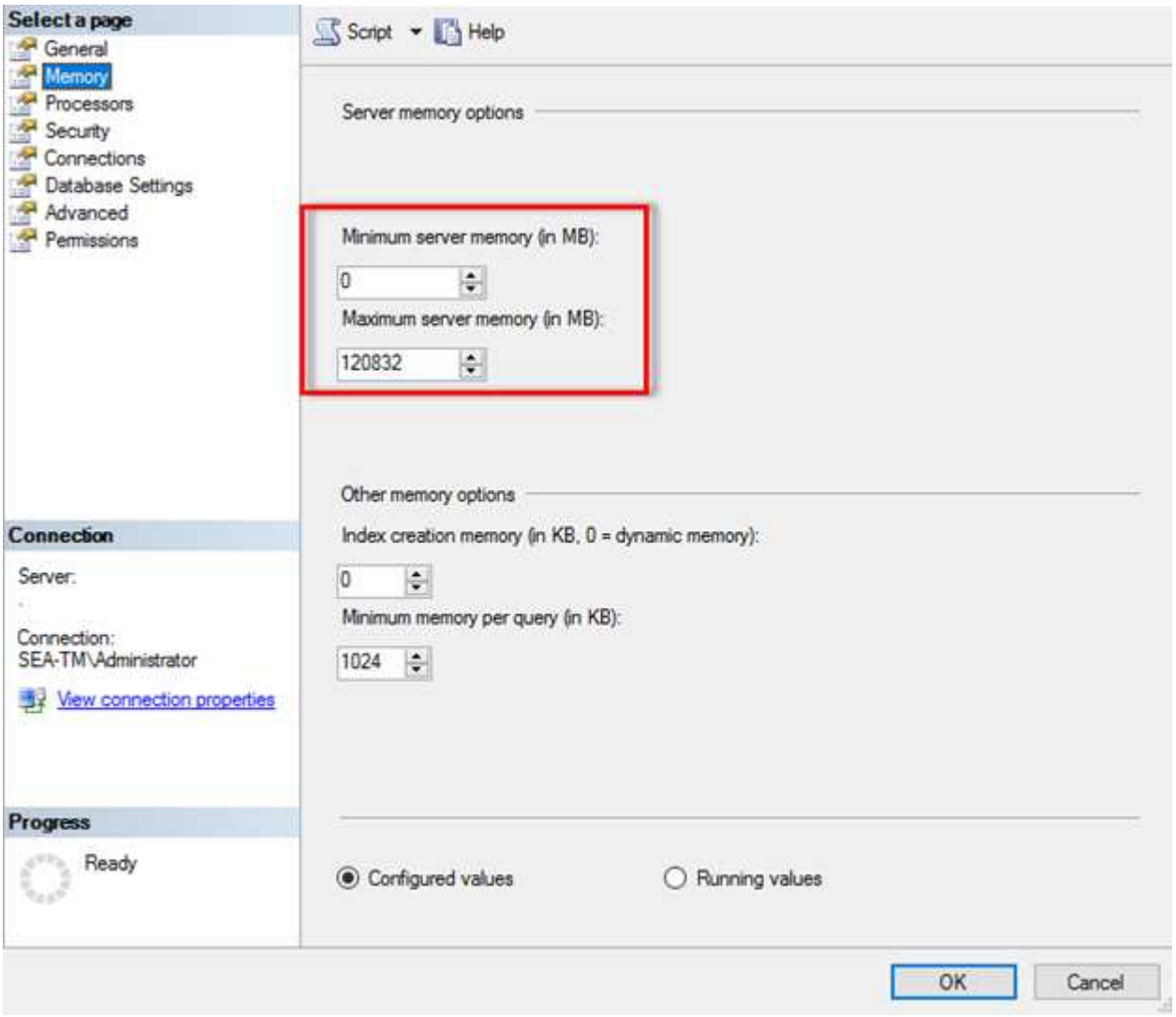
Mémoire maximale du serveur

L'option max. De mémoire du serveur définit la quantité maximale de mémoire que l'instance SQL Server peut utiliser. Il est généralement utilisé si plusieurs applications s'exécutent sur le même serveur que SQL Server et que vous voulez vous assurer que ces applications disposent de suffisamment de mémoire pour fonctionner correctement.

Certaines applications n'utilisent que la mémoire disponible au démarrage et ne demandent pas de mémoire supplémentaire, même si elles sont sous pression. C'est là que le paramètre de mémoire maximale du serveur entre en jeu.

Sur un cluster SQL Server avec plusieurs instances SQL Server, chaque instance peut être en concurrence pour des ressources. La définition d'une limite de mémoire pour chaque instance de SQL Server peut aider à garantir les meilleures performances pour chaque instance.

NetApp recommande de laisser au moins 4 Go à 6 Go de RAM pour le système d'exploitation afin d'éviter les problèmes de performances.



Réglage de la mémoire minimale et maximale du serveur à l'aide de SQL Server Management Studio.

L'utilisation de SQL Server Management Studio pour ajuster la mémoire minimale ou maximale du serveur nécessite un redémarrage du service SQL Server. Vous pouvez également ajuster la mémoire du serveur à

l'aide de Trantransaction SQL (T-SQL) à l'aide du code suivant :

```
EXECUTE sp_configure 'show advanced options', 1
GO
EXECUTE sp_configure 'min server memory (MB)', 2048
GO
EXEC sp_configure 'max server memory (MB)', 120832
GO
RECONFIGURE WITH OVERRIDE
```

Accès à la mémoire non uniforme

L'accès à la mémoire non uniforme (NUMA) est une technologie d'optimisation de l'accès à la mémoire qui permet d'éviter une charge supplémentaire sur le bus du processeur.

Si NUMA est configuré sur un serveur sur lequel SQL Server est installé, aucune configuration supplémentaire n'est requise car SQL Server est compatible avec NUMA et fonctionne bien sur le matériel NUMA.

Mémoire de création d'index

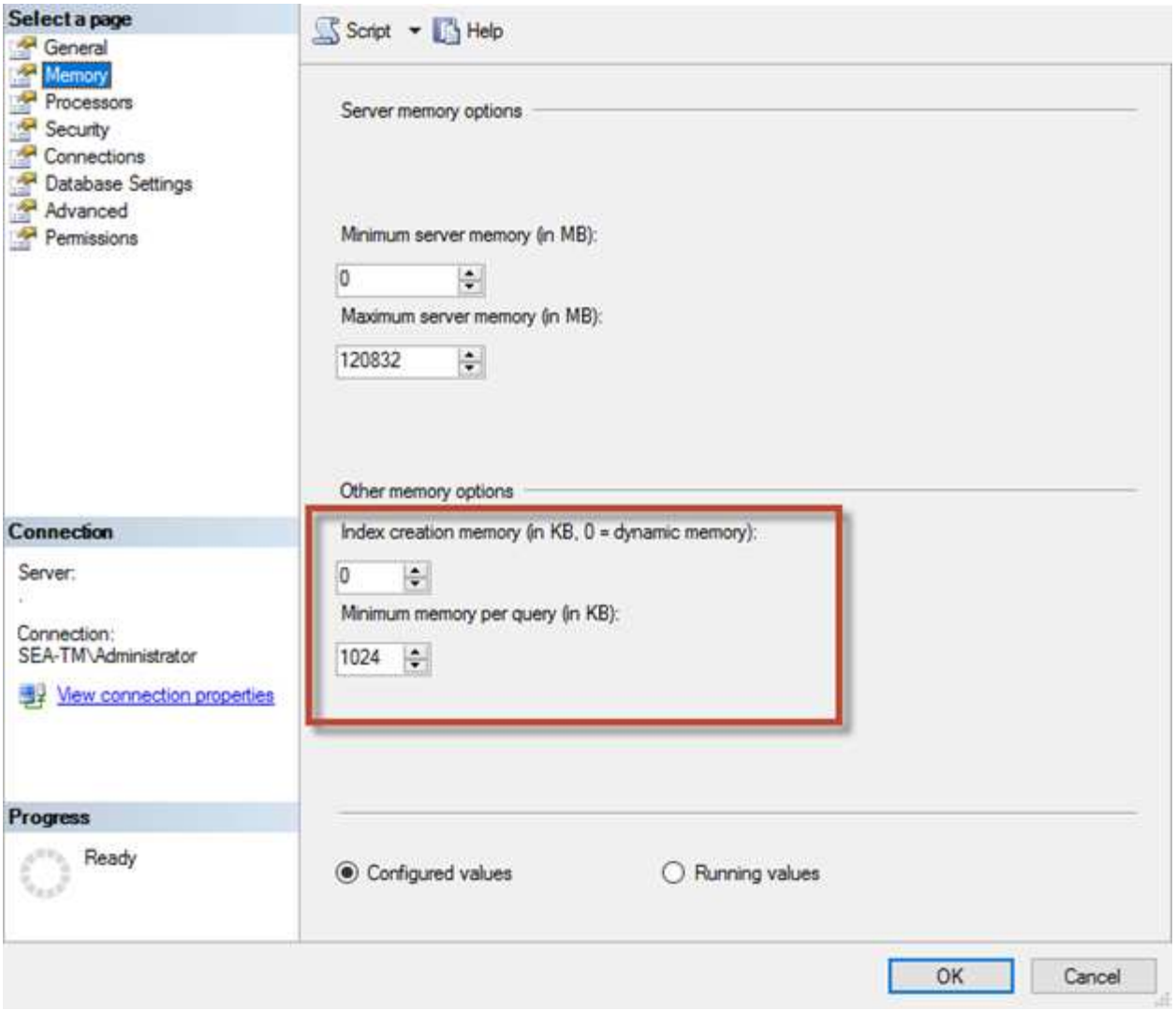
L'option index create memory est une autre option avancée qui ne devrait normalement pas avoir besoin d'être modifiée par défaut.

Il contrôle la quantité maximale de RAM initialement allouée pour la création d'index. La valeur par défaut de cette option est 0, ce qui signifie qu'elle est gérée automatiquement par SQL Server. Cependant, si vous rencontrez des difficultés à créer des index, envisagez d'augmenter la valeur de cette option.

Mémoire min. Par requête

Lorsqu'une requête est exécutée, SQL Server tente d'allouer la quantité optimale de mémoire pour s'exécuter efficacement.

Par défaut, la mémoire min par paramètre de requête alloue >= à 1024 Ko pour chaque requête à exécuter. Il est recommandé de laisser ce paramètre à la valeur par défaut afin de permettre à SQL Server de gérer de façon dynamique la quantité de mémoire allouée aux opérations de création d'index. Cependant, si SQL Server dispose de plus de RAM que nécessaire pour fonctionner efficacement, les performances de certaines requêtes peuvent être améliorées si vous augmentez ce paramètre. Par conséquent, tant que la mémoire est disponible sur le serveur qui n'est pas utilisé par SQL Server, toute autre application ou le système d'exploitation, l'augmentation de ce paramètre peut aider à améliorer les performances globales de SQL Server. Si aucune mémoire disponible n'est disponible, l'augmentation de ce paramètre peut nuire aux performances globales.



Instance partagée ou instance dédiée

SQL Server peut être configuré en tant qu'instance unique par serveur ou en tant que plusieurs instances. La bonne décision dépend généralement de facteurs tels que l'utilisation du serveur pour la production ou le développement, que l'instance soit considérée comme stratégique pour le fonctionnement de l'entreprise et les objectifs de performances.

Les configurations d'instances partagées peuvent être initialement plus faciles à configurer, mais elles peuvent entraîner des problèmes de division ou de verrouillage des ressources, ce qui entraîne des problèmes de performances pour d'autres applications dont les bases de données sont hébergées sur l'instance SQL Server partagée.

La résolution des problèmes de performances peut s'avérer complexe, car vous devez déterminer quelle instance est la cause première. Cette question est comparée aux coûts des licences de systèmes d'exploitation et des licences SQL Server. Si les performances des applications sont primordiales, une instance dédiée est fortement recommandée.

Microsoft octroie des licences SQL Server par cœur au niveau du serveur et non par instance. C'est pourquoi les administrateurs de base de données sont tentés d'installer autant d'instances SQL Server que le serveur peut gérer pour réduire les coûts de licence, ce qui peut entraîner des problèmes de performances majeurs

par la suite.

NetApp recommande de choisir des instances SQL Server dédiées chaque fois que possible pour obtenir des performances optimales.

Fichiers tempdb

La base de données tempdb peut être largement utilisée. Outre le placement optimal des fichiers de base de données utilisateur sur ONTAP, le placement des fichiers de données tempdb est également essentiel pour réduire les conflits d'allocation. Tempdb doit être placé sur un disque distinct et non partagé avec les fichiers de données utilisateur.

Des conflits de page peuvent se produire sur les pages GAM (Global allocation map), SGAM (Shared global allocation map) ou PFS (page Free Space) lorsque SQL Server doit écrire sur des pages système spéciales pour allouer de nouveaux objets. Les loquets verrouillent ces pages en mémoire. Sur une instance SQL Server occupée, l'obtention d'un verrou sur une page système dans tempdb peut prendre un certain temps. Cela ralentit les temps d'exécution des requêtes et est appelé conflit de type LATCH. Consultez les meilleures pratiques suivantes pour la création de fichiers de données tempdb :

- Pour < ou = jusqu'à 8 cœurs : fichiers de données tempdb = nombre de cœurs
- Pour plus de 8 cœurs : 8 fichiers de données tempdb
- Le fichier de données tempdb doit être créé avec la même taille

L'exemple de script suivant modifie tempdb en créant huit fichiers tempdb de taille égale et en déplaçant tempdb vers le point de montage `C:\MSSQL\tempdb` pour SQL Server 2012 et versions ultérieures.

```
use master

go

-- Change logical tempdb file name first since SQL Server shipped with
logical file name called tempdev

alter database tempdb modify file (name = 'tempdev', newname =
'tempdev01');



-- Change location of tempdev01 and log file

alter database tempdb modify file (name = 'tempdev01', filename =
'C:\MSSQL\tempdb\tempdev01.mdf');

alter database tempdb modify file (name = 'templog', filename =
'C:\MSSQL\tempdb\templog.ldf');

GO
```

```
-- Assign proper size for tempdev01

ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'tempdev01', SIZE = 10GB );

ALTER DATABASE [tempdb] MODIFY FILE ( NAME = N'templog', SIZE = 10GB );

GO

-- Add more tempdb files

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev02', FILENAME =
N'C:\MSSQL\tempdb\tempdev02.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev03', FILENAME =
N'C:\MSSQL\tempdb\tempdev03.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev04', FILENAME =
N'C:\MSSQL\tempdb\tempdev04.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev05', FILENAME =
N'C:\MSSQL\tempdb\tempdev05.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev06', FILENAME =
N'C:\MSSQL\tempdb\tempdev06.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev07', FILENAME =
N'C:\MSSQL\tempdb\tempdev07.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

ALTER DATABASE [tempdb] ADD FILE ( NAME = N'tempdev08', FILENAME =
N'C:\MSSQL\tempdb\tempdev08.ndf' , SIZE = 10GB , FILEGROWTH = 10%);

GO
```

À partir de SQL Server 2016, le nombre de cœurs de CPU visibles par le système d'exploitation est automatiquement détecté lors de l'installation et, en fonction de ce nombre, SQL Server calcule et configure le nombre de fichiers tempdb requis pour des performances optimales.

Configuration du stockage sur les systèmes AFF/FAS

Présentation

L'association des solutions de stockage ONTAP et de Microsoft SQL Server permet de concevoir un stockage de base de données adapté aux besoins des applications les plus exigeantes.

Pour optimiser une solution SQL Server sur ONTAP, il est nécessaire de comprendre le modèle et les caractéristiques d'E/S SQL Server. Une infrastructure de stockage bien conçue pour une base de données

SQL Server doit répondre aux besoins de performances de SQL Server, tout en assurant une gestion maximale de l'infrastructure dans son ensemble. Une bonne disposition du stockage permet également de réussir le déploiement initial et d'assurer une croissance progressive de l'environnement à mesure que l'entreprise se développe.

Conception du stockage des données

Pour les bases de données SQL Server qui n'utilisent pas SnapCenter pour effectuer des sauvegardes, Microsoft recommande de placer les données et les fichiers journaux sur des disques distincts. Pour les applications qui mettent à jour et demandent simultanément des données, le fichier journal est très gourmand en écriture et le fichier de données (selon votre application) consomme beaucoup de ressources en lecture/écriture. Pour la récupération des données, le fichier journal n'est pas nécessaire. Par conséquent, les demandes de données peuvent être satisfaites à partir du fichier de données placé sur son propre disque.

Lorsque vous créez une nouvelle base de données, Microsoft recommande de spécifier des disques distincts pour les données et les journaux. Pour déplacer des fichiers après la création de la base de données, la base de données doit être mise hors ligne. Pour plus d'informations sur les recommandations de Microsoft, consultez la section "Placez les fichiers de données et les fichiers journaux sur des lecteurs distincts".

64 bits

Les agrégats constituent les conteneurs de stockage de niveau le plus bas pour les configurations de stockage NetApp. Il existe sur Internet une documentation existante qui recommande de séparer les E/S sur différents jeux de disques sous-jacents. Ceci n'est pas recommandé avec ONTAP. NetApp a effectué plusieurs tests de caractérisation des charges de travail d'E/S à l'aide d'agrégats partagés et dédiés, avec des fichiers de données et des fichiers journaux de transactions séparés. Les tests montrent qu'un grand agrégat avec plus de disques et de groupes RAID optimise et améliore les performances du stockage et est plus facile à gérer pour les administrateurs pour deux raisons :

- Un grand agrégat rend les capacités d'E/S de tous les disques disponibles pour tous les fichiers.
- Un seul grand agrégat permet d'optimiser l'utilisation de l'espace disque.

Pour la haute disponibilité (HA), placer la réplique synchrone secondaire SQL Server Always On Availability Group sur une machine virtuelle de stockage (SVM) distincte dans l'agrégat. Pour la reprise sur incident, placez la réplication asynchrone sur un agrégat faisant partie d'un cluster de stockage distinct dans le site de reprise sur incident, le contenu étant répliqué à l'aide de la technologie NetApp SnapMirror. Pour des performances de stockage optimales, NetApp recommande de disposer d'au moins 10 % d'espace libre dans un agrégat.

Volumes

les volumes sont créés et résident dans des agrégats. Ce terme peut parfois engendrer une confusion, car un volume ONTAP n'est pas une LUN. Un volume ONTAP est un conteneur de gestion de données. Un volume peut contenir des fichiers, des LUN ou même des objets S3. Un volume ne prend pas d'espace, il est uniquement utilisé pour la gestion des données contenues.

Considérations relatives à la conception des volumes

Avant de créer une conception de volume de base de données, il est important de comprendre comment le modèle et les caractéristiques d'E/S SQL Server varient en fonction de la charge de travail et des exigences de sauvegarde et de restauration. Consultez les recommandations NetApp suivantes pour les volumes flexibles :

- Évitez de partager des volumes entre des hôtes. Par exemple, s'il est possible de créer 2 LUN dans un

seul volume et de partager chaque LUN avec un autre hôte, cela peut être évité, car la gestion peut en compliquer la tâche. Dans le cas d'une exécution de plusieurs instances SQL Server sur le même hôte, sauf si vous êtes proche de la limite de volume sur un nœud, évitez le partage de volume et disposez à la place d'un volume distinct par instance et par hôte pour faciliter la gestion des données.

- Utilisez des points de montage NTFS au lieu de lettres de lecteur pour dépasser la limite de 26 lettres de lecteur dans Windows. Lorsque vous utilisez des points de montage de volume, il est généralement recommandé de donner au libellé de volume le même nom que le point de montage.
- Le cas échéant, configurez une règle de dimensionnement automatique de volume pour éviter les conditions de manque d'espace.
- Si vous installez SQL Server sur un partage SMB, assurez-vous que Unicode est activé sur les volumes SMB pour la création de dossiers.
- Définissez la valeur de la réserve d'instantanés dans le volume sur zéro pour faciliter la surveillance d'un point de vue opérationnel.
- Désactivez les planifications d'instantanés et les stratégies de conservation. Utilisez plutôt SnapCenter pour coordonner les copies Snapshot des volumes de données SQL Server.
- Placez les bases de données système SQL Server sur un volume dédié.
- Tempdb est une base de données système utilisée par SQL Server comme espace de travail temporaire, en particulier pour les opérations DBCC CHECKDB exigeantes en E/S. Par conséquent, placez cette base de données sur un volume dédié avec un jeu séparé de piles de disques. Dans les grands environnements dans lesquels le nombre de volumes est un défi, vous pouvez consolider tempdb en un nombre réduit de volumes et le stocker dans le même volume que les autres bases de données système après une planification minutieuse. La protection des données pour tempdb n'est pas une priorité élevée car cette base de données est recréée à chaque redémarrage de SQL Server.
- Placez les fichiers de données utilisateur (`.mdf`) sur des volumes distincts, car il s'agit de charges de travail en lecture/écriture aléatoires. Il est courant de créer des sauvegardes du journal de transactions plus fréquemment que les sauvegardes de bases de données. Pour cette raison, placez les fichiers journaux (`.ldf`de transactions) sur un volume distinct ou un fichier VMDK à partir des fichiers de données afin que des planifications de sauvegarde indépendantes puissent être créées pour chaque. Cette séparation isole également les E/S d'écriture séquentielle des fichiers journaux des E/S de lecture/écriture aléatoires des fichiers de données et améliore considérablement les performances de SQL Server.

LUN

- Assurez-vous que les fichiers de base de données utilisateur et le répertoire des journaux pour stocker la sauvegarde des journaux se trouvent sur des volumes distincts afin d'empêcher la règle de conservation d'écraser les snapshots lorsqu'ils sont utilisés avec la technologie SnapVault.
- Ne mélangez pas des fichiers de base de données et des fichiers autres que des bases de données, tels que des fichiers de recherche en texte intégral, sur le même LUN.
- Le placement de fichiers secondaires de base de données (dans le cadre d'un groupe de fichiers) sur des volumes distincts améliore les performances de la base de données SQL Server. Cette séparation est valide uniquement si le fichier de la base de données `.mdf` ne partage pas son LUN avec d'autres `.mdf` fichiers.
- Si vous créez des LUN à l'aide de DiskManager ou d'autres outils, assurez-vous que la taille de l'unité d'allocation est définie sur 64 Ko pour les partitions lors du formatage des LUN.
- Voir la "Microsoft Windows et MPIO natif conformément aux meilleures pratiques ONTAP pour les SAN modernes" Pour appliquer la prise en charge des chemins d'accès multiples sur Windows aux périphériques iSCSI dans les propriétés MPIO.

Fichiers de base de données et groupes de fichiers

Il est essentiel de placer correctement les fichiers de base de données SQL Server sur ONTAP lors de la phase de déploiement initiale. Vous bénéficiez ainsi de performances optimales, d'un temps de gestion de l'espace, de sauvegarde et de restauration qui peuvent être configurés pour répondre aux besoins de votre entreprise.

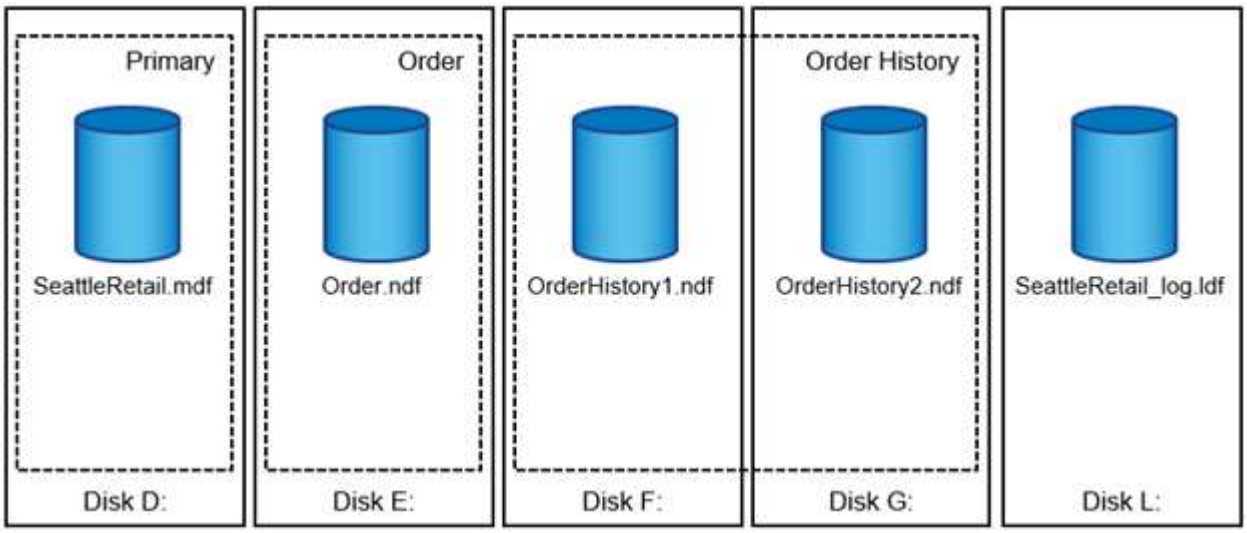
En théorie, SQL Server (64 bits) prend en charge 32,767 bases de données par instance et 524 272 To de taille de base de données, bien que l'installation standard comporte généralement plusieurs bases de données. Cependant, le nombre de bases de données que SQL Server peut gérer dépend de la charge et du matériel. Il n'est pas rare que des instances SQL Server hébergent des dizaines, des centaines, voire des milliers de petites bases de données.

Fichiers de base de données et groupe de fichiers

Chaque base de données se compose d'un ou plusieurs fichiers de données et d'un ou plusieurs fichiers journaux de transactions. Le journal de transactions stocke les informations sur les transactions de base de données et toutes les modifications de données effectuées par chaque session. Chaque fois que les données sont modifiées, SQL Server stocke suffisamment d'informations dans le journal de transactions pour annuler (revenir en arrière) ou rétablir (relire) l'action. Un journal de transactions SQL Server fait partie intégrante de la réputation de SQL Server en matière d'intégrité et de robustesse des données. Le journal de transactions est essentiel aux capacités d'atomicité, de cohérence, d'isolation et de durabilité (ACIDE) de SQL Server. SQL Server écrit dans le journal de transactions dès qu'une modification de la page de données se produit. Chaque instruction Data manipulation Language (DML) (par exemple, Select, INSERT, Update ou DELETE) est une transaction complète, et le journal de transactions s'assure que l'opération basée sur l'ensemble a lieu, en s'assurant de l'atomicité de la transaction.

Chaque base de données possède un fichier de données primaire, qui, par défaut, possède l'extension .mdf. En outre, chaque base de données peut avoir des fichiers de base de données secondaires. Ces fichiers, par défaut, ont des extensions .ndf.

Tous les fichiers de base de données sont regroupés en groupes de fichiers. Un groupe de fichiers est l'unité logique, qui simplifie l'administration de la base de données. Ils permettent de séparer le placement d'objets logiques des fichiers de base de données physiques. Lorsque vous créez les tables d'objets de base de données, vous spécifiez dans quel groupe de fichiers elles doivent être placées sans vous soucier de la configuration du fichier de données sous-jacent.



La possibilité de placer plusieurs fichiers de données dans le groupe de fichiers vous permet de répartir la charge entre les différents périphériques de stockage, ce qui contribue à améliorer les performances d'E/S du système. En revanche, le journal de transactions ne bénéficie pas des multiples fichiers car SQL Server écrit dans le journal de transactions de manière séquentielle.

La séparation entre le placement d'objets logiques dans les groupes de fichiers et les fichiers de base de données physiques vous permet d'affiner la disposition des fichiers de base de données, en tirant le meilleur parti du sous-système de stockage. Le nombre de fichiers de données prenant en charge une charge de travail donnée peut varier en fonction des besoins pour prendre en charge les exigences d'E/S et la capacité prévue, sans affecter l'application. Ces variations dans la disposition de la base de données sont transparentes pour les développeurs d'applications, qui placent les objets de base de données dans les groupes de fichiers plutôt que dans les fichiers de base de données.

NetApp recommande d'éviter l'utilisation du groupe de fichiers principal pour tout autre objet que les objets système. La création d'un groupe de fichiers distinct ou d'un ensemble de groupes de fichiers pour les objets utilisateur simplifie l'administration de la base de données et la reprise après incident, en particulier dans le cas de bases de données volumineuses.

Initialisation du fichier d'instance de base de données

Vous pouvez spécifier la taille initiale du fichier et les paramètres de croissance automatique au moment de la création de la base de données ou de l'ajout de nouveaux fichiers à une base de données existante. SQL Server utilise un algorithme de remplissage proportionnel lors du choix du fichier de données dans lequel il doit écrire des données. Elle écrit une quantité de données proportionnellement à l'espace libre disponible dans les fichiers. Plus l'espace libre dans le fichier est important, plus il traite d'écritures.

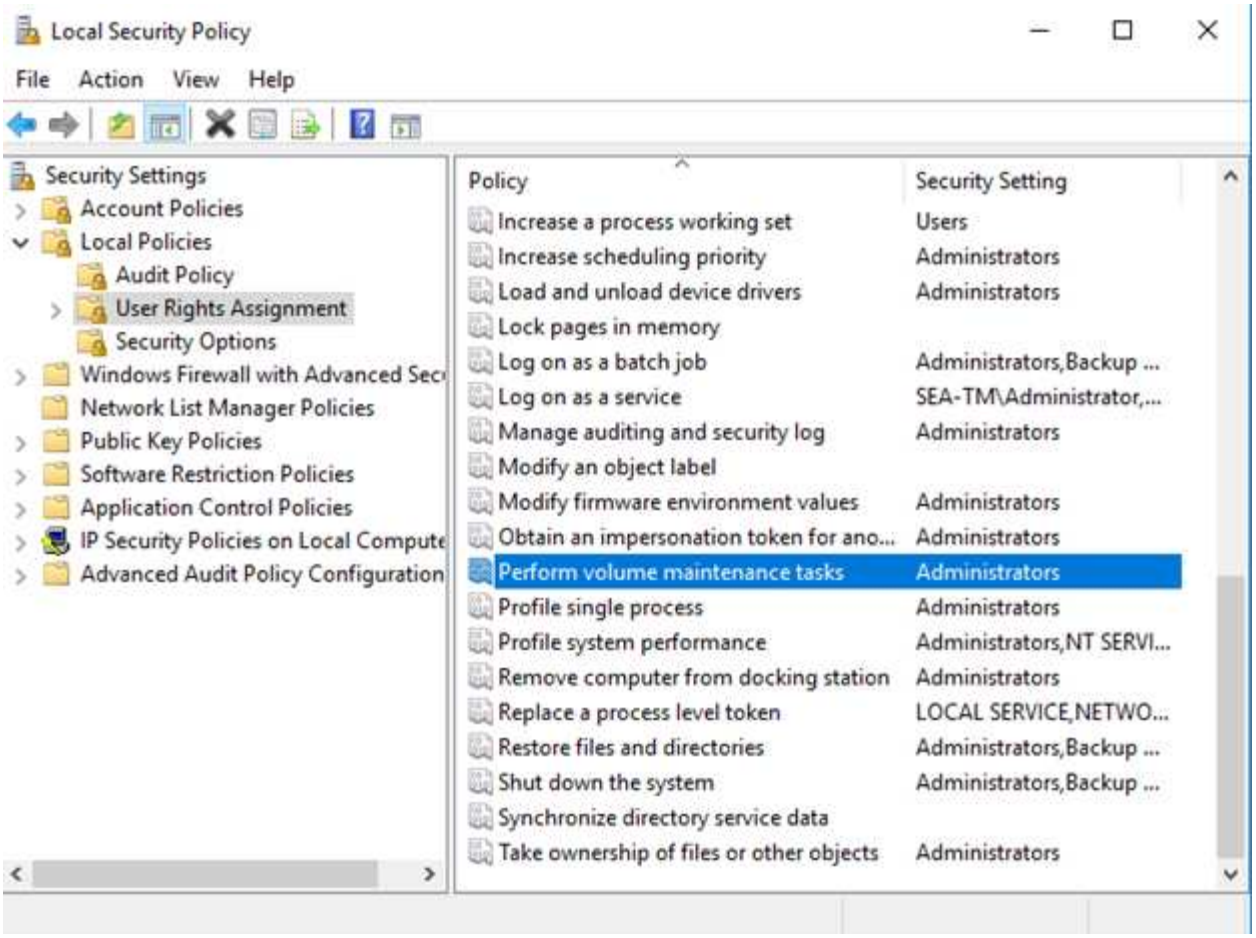
NetApp recommande que tous les fichiers d'un seul groupe de fichiers aient les mêmes paramètres de taille initiale et de croissance automatique, avec la taille de croissance définie en mégaoctets plutôt qu'en pourcentages. Cela permet à l'algorithme de remplissage proportionnel d'équilibrer uniformément les activités d'écriture entre les fichiers de données.

Chaque fois que SQL Server augmente la taille des fichiers, il remplit l'espace nouvellement alloué avec des zéros. Ce processus bloque toutes les sessions qui doivent écrire dans le fichier correspondant ou, en cas de croissance du journal de transactions, générer des enregistrements de journal de transactions.

SQL Server met toujours à zéro le journal de transactions et ce comportement ne peut pas être modifié. Toutefois, vous pouvez contrôler si les fichiers de données sont mis à zéro en activant ou en désactivant l'initialisation instantanée des fichiers. L'activation de l'initialisation instantanée des fichiers permet d'accélérer la croissance des fichiers de données et de réduire le temps nécessaire à la création ou à la restauration de la base de données.

Un petit risque de sécurité est associé à l'initialisation instantanée des fichiers. Lorsque cette option est activée, les parties non allouées du fichier de données peuvent contenir des informations provenant de fichiers OS précédemment supprimés. Les administrateurs de base de données peuvent examiner ces données.

Vous pouvez activer l'initialisation instantanée des fichiers en ajoutant l'autorisation sa_MANAGE_VOLUME_NAME, également appelée « effectuer une tâche de maintenance de volume » au compte de démarrage SQL Server. Vous pouvez le faire sous l'application de gestion des stratégies de sécurité locales (secpol.msc), comme indiqué dans la figure suivante. Ouvrez les propriétés de l'autorisation "effectuer une tâche de maintenance de volume" et ajoutez le compte de démarrage SQL Server à la liste des utilisateurs.



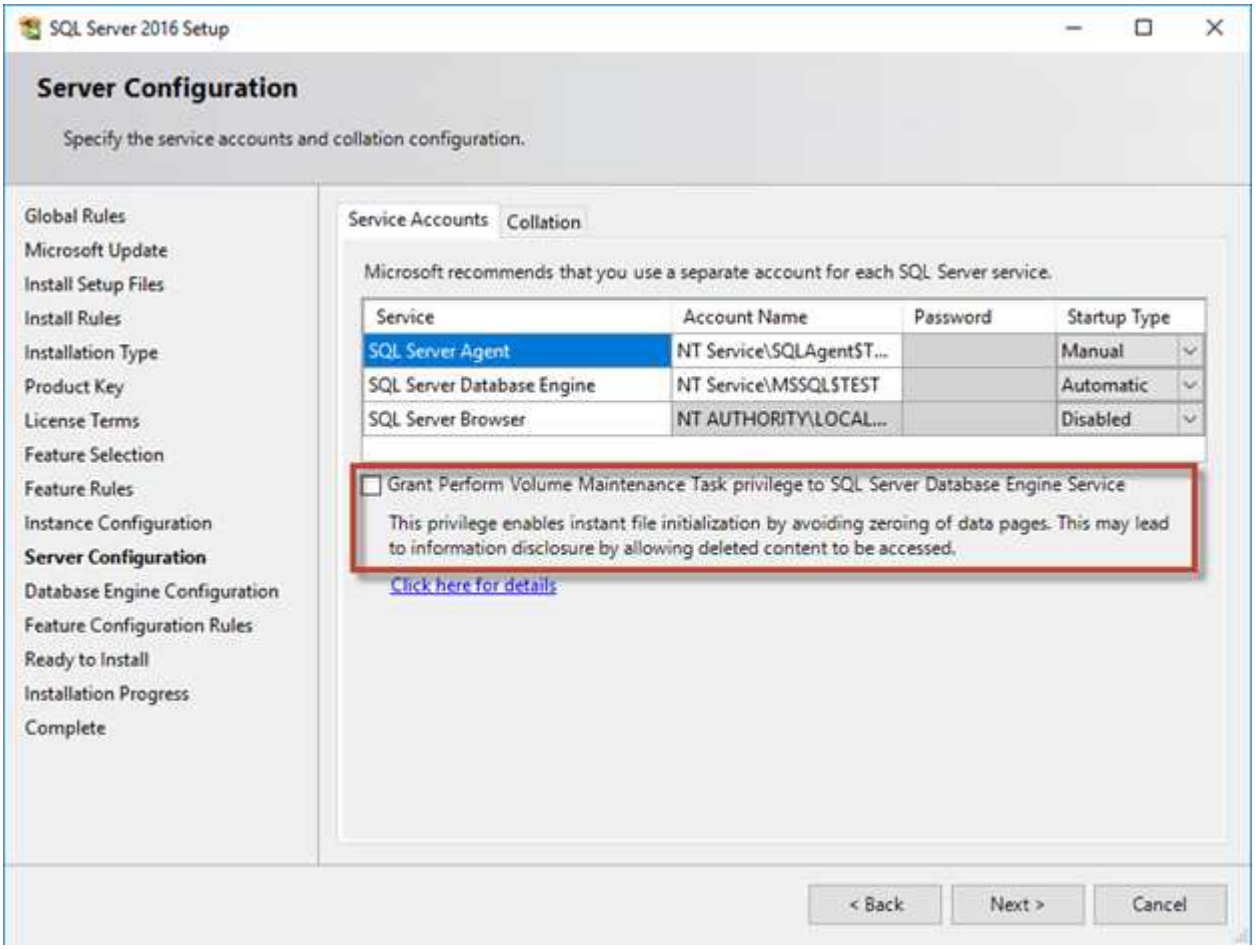
Pour vérifier si l'autorisation est activée, vous pouvez utiliser le code de l'exemple suivant. Ce code définit deux indicateurs de suivi qui forcent SQL Server à écrire des informations supplémentaires dans le journal d'erreurs, à créer une petite base de données et à lire le contenu du journal.

```
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO
```

Lorsque l'initialisation instantanée des fichiers n'est pas activée, le journal d'erreurs SQL Server indique que SQL Server met à zéro le fichier de données mdf en plus de mettre à zéro le fichier journal ldf, comme indiqué dans l'exemple suivant. Lorsque l'initialisation instantanée des fichiers est activée, elle affiche uniquement la remise à zéro du fichier journal.

	LogDate	ProcessInfo	Text
365	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 flush delta counts.
366	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 logging active xact info.
367	2017-02-09 08:10:07.750	spid53	Ckpt dbid 3 phase 1 ended (8)
368	2017-02-09 08:10:07.750	spid53	About to log Checkpoint end.
369	2017-02-09 08:10:07.880	spid53	Ckpt dbid 3 complete
370	2017-02-09 08:10:08.130	spid53	Starting up database 'DelMe'.
371	2017-02-09 08:10:08.150	spid53	FixupLog Tail(progress) zeroing C:\Program Files\Micros
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQ
373	2017-02-09 08:10:08.170	spid53	Zeroing completed on C:\Program Files\Microsoft SQL
374	2017-02-09 08:10:08.710	spid53	Ckpt dbid 6 started
375	2017-02-09 08:10:08.710	spid53	About to log Checkpoint begin.

La tâche effectuer une maintenance de volume est simplifiée dans SQL Server 2016 et est fournie ultérieurement en option pendant le processus d'installation. Cette figure affiche l'option permettant d'accorder au service du moteur de base de données SQL Server le privilège d'effectuer la tâche de maintenance du volume.



Une autre option de base de données importante qui contrôle la taille des fichiers de base de données est la fonction de transmission automatique. Lorsque cette option est activée, SQL Server réduit régulièrement les fichiers de base de données, réduit leur taille et libère de l'espace dans le système d'exploitation. Cette opération consomme beaucoup de ressources et est rarement utile car les fichiers de base de données augmentent à nouveau après l'arrivée de nouvelles données dans le système. La fonction Autohrink ne doit pas être activée sur la base de données.

Répertoire du journal

Le répertoire du journal est spécifié dans SQL Server pour stocker les données de sauvegarde du journal de transactions au niveau de l'hôte. Si vous utilisez SnapCenter pour sauvegarder les fichiers journaux, chaque hôte SQL Server utilisé par SnapCenter doit disposer d'un répertoire de journaux hôte configuré pour effectuer des sauvegardes de journaux. SnapCenter dispose d'un référentiel de base de données. Les métadonnées liées aux opérations de sauvegarde, de restauration ou de clonage sont donc stockées dans un référentiel de base de données central.

La taille du répertoire de journaux hôte est calculée comme suit :
Taille du répertoire des journaux hôtes = ((taille LDF maximale de la base de données x taux de modification quotidien du journal %) x (rétention des snapshots) ÷ (1 - espace de surcharge de la LUN %)
La formule de dimensionnement du répertoire des journaux hôte suppose un espace supplémentaire de 10 % pour les LUN

Placez le répertoire des journaux sur un volume ou une LUN dédié. La quantité de données dans le répertoire du journal hôte dépend de la taille des sauvegardes et du nombre de jours pendant lesquels les sauvegardes sont conservées. SnapCenter n'autorise qu'un seul répertoire de journaux hôte par hôte SQL Server. Vous pouvez configurer les répertoires de journaux hôtes dans SnapCenter -→ hôte -→ configurer le plug-in.

NetApp recommande ce qui suit pour un répertoire de journaux hôte :

- Assurez-vous que le répertoire du journal de l'hôte n'est partagé par aucun autre type de données pouvant potentiellement corrompre les données du snapshot de sauvegarde.
- Ne placez pas de bases de données utilisateur ou de bases de données système sur un LUN qui héberge des points de montage.
- Créez le répertoire des journaux hôtes sur un volume dédié sur lequel SnapCenter copie les journaux de transactions.
- Utilisez les assistants SnapCenter pour migrer les bases de données vers le stockage NetApp de sorte que les bases de données soient stockées dans des emplacements valides, ce qui permet de réaliser les opérations de sauvegarde et de restauration SnapCenter. N'oubliez pas que le processus de migration est disruptif et peut mettre les bases de données hors ligne pendant la migration.
- Les conditions suivantes doivent être en place pour les instances de cluster de basculement (FCI) de SQL Server :
  - Si vous utilisez une instance de cluster de basculement, la LUN du répertoire de journalisation de l'hôte doit être une ressource de disque de cluster dans le même groupe de cluster que l'instance SQL Server en cours de sauvegarde SnapCenter.
  - Si vous utilisez une instance de cluster de basculement, les bases de données utilisateur doivent être placées sur des LUN partagées qui sont des ressources de cluster de disques physiques affectées au groupe de clusters associé à l'instance SQL Server.

Efficacité du stockage

L'efficacité du stockage ONTAP est optimisée pour stocker et gérer les données SQL Server de manière à utiliser la quantité la plus faible d'espace de stockage sans affecter les performances.

Les fonctionnalités d'optimisation de l'espace, telles que la compression, la compaction et la déduplication, sont conçues pour augmenter la quantité de données logiques correspondant à un volume de stockage

physique donné. Vous réduisez ainsi vos coûts et vos frais de gestion.

À un niveau élevé, la compression est un processus mathématique qui permet de détecter et d'encoder des modèles de données de manière à réduire les besoins en espace. En revanche, la déduplication détecte les blocs de données répétés et supprime les copies parasites. La compaction permet à plusieurs blocs logiques de données de partager le même bloc physique sur le support.

> Reportez-vous aux sections ci-dessous sur le provisionnement fin pour une explication de l'interaction entre l'efficacité du stockage et la réservation fractionnaire.

Compression

Avant la disponibilité des systèmes de stockage 100 % Flash, la compression basée sur les baies était d'une valeur limitée, car la plupart des charges de travail exigeantes en E/S nécessitaient un très grand nombre de piles pour obtenir une performance acceptable. Les systèmes de stockage contenaient invariablement beaucoup plus de capacité que nécessaire, ce qui a pour effet d'augmenter le nombre de disques. La situation a changé avec la montée du stockage Solid-State. Il n'est plus nécessaire de surprovisionner des disques uniquement pour obtenir de bonnes performances. L'espace disque d'un système de stockage peut être adapté aux besoins réels en termes de capacité.

La capacité accrue des disques SSD en termes d'IOPS permet presque toujours de réaliser des économies par rapport aux disques rotatifs. Toutefois, la compression peut réaliser davantage d'économies en augmentant la capacité effective des supports SSD.

Il existe plusieurs façons de compresser les données. De nombreuses bases de données incluent leurs propres fonctionnalités de compression, mais ce phénomène est rarement observé dans les environnements clients. La raison en est généralement la réduction des performances pour un **changement** de données compressées, plus avec certaines applications, il existe des coûts de licence élevés pour la compression au niveau de la base de données. Enfin, il y a les conséquences globales sur les performances des opérations des bases de données. Il est peu judicieux de payer un coût de licence par processeur élevé pour un processeur qui effectue la compression et la décompression des données plutôt que le véritable travail de base de données. Une meilleure option consiste à décharger la tâche de compression sur le système de stockage.

Compression adaptative

La compression adaptative a été testée en profondeur avec des charges de travail exigeantes sans effet sur les performances, même dans un environnement 100 % Flash où la latence se mesure en microsecondes. Certains clients ont même signalé une augmentation des performances due à l'utilisation de la compression, car les données restent compressées dans le cache, augmentant ainsi la quantité de cache disponible dans un contrôleur.

ONTAP gère les blocs physiques dans des unités de 4 Ko. La compression adaptative utilise une taille de bloc de compression par défaut de 8 Ko, ce qui signifie que les données sont compressées dans des unités de 8 Ko. La taille de bloc de 8 Ko la plus utilisée par les bases de données relationnelles est donc identique. Les algorithmes de compression deviennent plus efficaces avec la compression d'un volume croissant de données. Une taille de bloc de compression de 32 Ko serait plus compacte qu'une unité de bloc de compression de 8 Ko. Cela signifie que la compression adaptative utilisant une taille de bloc de 8 Ko par défaut entraîne des taux d'efficacité légèrement inférieurs, mais qu'une taille de bloc de compression inférieure présente également des avantages considérables. Les charges de travail de la base de données incluent une grande quantité d'activités de remplacement. Le remplacement d'un bloc de données de 32 Ko compressé de 8 Ko nécessite la lecture de l'intégralité des 32 Ko de données logiques, leur décompression, la mise à jour de la région de 8 Ko requise, la recompression, puis l'écriture de la totalité des 32 Ko sur les disques. Cette opération est très coûteuse pour un système de stockage. En effet, certaines baies de stockage concurrentes, basées sur des

blocs de compression plus volumineux, affectent également considérablement les performances des charges de travail de la base de données.

> La taille de bloc utilisée par la compression adaptative peut être augmentée jusqu'à 32 Ko. Cela peut améliorer l'efficacité du stockage et doit être envisagé pour les fichiers de repos tels que les journaux de transactions et les fichiers de sauvegarde lorsqu'une quantité importante de ces données est stockée sur la baie. Dans certains cas, les bases de données actives qui utilisent une taille de bloc de 16 ou 32 Ko peuvent également tirer parti de l'augmentation de la taille de bloc de la compression adaptative pour qu'elle corresponde. Consultez un représentant NetApp ou partenaire pour savoir si cette solution convient à votre charge de travail.

> Les tailles de bloc de compression supérieures à 8 Ko ne doivent pas être utilisées avec la déduplication sur les destinations de sauvegarde en streaming. Les petites modifications apportées aux données sauvegardées affectent la fenêtre de compression de 32 Ko. Si la fenêtre change, les données compressées obtenues diffèrent dans l'ensemble du fichier. La déduplication a lieu après la compression, ce qui signifie que le moteur de déduplication voit chaque sauvegarde compressée différemment. Si la déduplication des sauvegardes en continu est nécessaire, seule une compression adaptative de bloc de 8 Ko doit être utilisée. Il est préférable d'utiliser la compression adaptative, car elle fonctionne à des blocs de taille réduite sans perturber l'efficacité de la déduplication. Pour des raisons similaires, la compression côté hôte interfère également avec l'efficacité de la déduplication.

Alignement de compression

La compression adaptative dans un environnement de base de données nécessite un certain respect de l'alignement des blocs de compression. Cela ne préoccupe que les données soumises à des écrasements aléatoires de blocs très spécifiques. Cette approche est similaire à l'alignement global du système de fichiers, où le début d'un système de fichiers doit être aligné sur une limite de périphérique de 4 Ko et la taille de bloc d'un système de fichiers doit être un multiple de 4 Ko.

Par exemple, une écriture de 8 Ko dans un fichier est compressée uniquement si elle s'aligne sur une limite de 8 Ko dans le système de fichiers lui-même. Ce point signifie qu'il doit figurer sur le premier 8 Ko du fichier, le deuxième 8 Ko du fichier, etc. La manière la plus simple de garantir un alignement correct est d'utiliser le type de LUN correct, toute partition créée doit avoir un décalage par rapport au début du périphérique qui est un multiple de 8K, et utiliser une taille de bloc du système de fichiers qui est un multiple de la taille de bloc de la base de données.

Les données telles que les sauvegardes ou les journaux de transactions sont des opérations écrites de manière séquentielle sur plusieurs blocs, qui sont tous compressés. Par conséquent, il n'est pas nécessaire de considérer l'alignement. Le seul modèle d'E/S préoccupant est l'écrasement aléatoire des fichiers.

Compaction

La compaction est une technologie qui améliore l'efficacité de la compression. Comme indiqué précédemment, la compression adaptative à elle seule permet d'économiser 2:1 au maximum, car elle se limite au stockage d'une E/S de 8 Ko dans un bloc WAFL de 4 Ko. Les méthodes de compression avec des blocs de taille supérieure améliorent l'efficacité. Cependant, elles ne conviennent pas aux données soumises à des remplacements de blocs de petite taille. La décompression d'unités de données de 32 Ko, la mise à jour d'une partie de 8 Ko, la recompression et l'écriture sur les disques entraînent une surcharge.

La compaction des données permet de stocker plusieurs blocs logiques dans des blocs physiques. Par exemple, une base de données avec des données fortement compressibles comme des blocs texte ou partiellement pleins peut être compressée de 8 Ko à 1 Ko. Sans compaction, 1 Ko de données occuperaient toujours un bloc complet de 4 Ko. La compaction des données à la volée permet de stocker 1 Ko de données

compressées dans un espace physique de seulement 1 Ko, parallèlement à d'autres données compressées. Il ne s'agit pas d'une technologie de compression. Il s'agit simplement d'un moyen plus efficace d'allouer de l'espace sur les disques et, par conséquent, il ne doit pas créer d'effet détectable sur les performances.

Le degré d'économie obtenu varie. En général, les données déjà compressées ou chiffrées ne peuvent pas être compressées davantage et, par conséquent, la compaction de ces datasets ne peut pas être bénéfique. À contrario, les fichiers de données récemment initialisés ne contiennent qu'un petit peu plus que des métadonnées de bloc et des zéros compressent jusqu'à 80:1.

Efficacité du stockage sensible à la température

L'efficacité de stockage sensible à la température (TSSE) est disponible dans ONTAP 9.8 et versions ultérieures. Il s'appuie sur des cartes thermiques d'accès aux blocs pour identifier les blocs peu utilisés et les compresser à l'aide d'une meilleure efficacité.

Déduplication

La déduplication permet de supprimer les tailles de bloc dupliquées d'un dataset. Par exemple, si le même bloc de 4 Ko existe dans 10 fichiers différents, la déduplication redirige ce bloc de 4 Ko au sein des 10 fichiers vers le même bloc physique de 4 Ko. Résultat : une amélioration de l'efficacité de ces données de 10:1.

Les données, telles que les LUN de démarrage invité VMware, se dédupliquent extrêmement bien, car elles sont constituées de plusieurs copies des mêmes fichiers du système d'exploitation. L'efficacité de 100:1 et plus ont été observées.

Certaines données ne contiennent pas de données dupliquées. Par exemple, un bloc Oracle contient un en-tête globalement unique à la base de données et une bande-annonce presque unique. Par conséquent, la déduplication d'une base de données Oracle permet rarement de réaliser plus de 1 % d'économies. La déduplication avec les bases de données MS SQL est légèrement meilleure, mais les métadonnées uniques au niveau des blocs restent une limitation.

Dans quelques cas, des économies d'espace allant jusqu'à 15 % ont été observées pour les bases de données de 16 Ko et les blocs volumineux. La bande de 4 Ko initiale de chaque bloc contient l'en-tête unique dans le monde, et le bloc de 4 Ko final contient la remorque presque unique. Les blocs internes sont candidats à la déduplication, bien que dans la pratique cela soit presque entièrement attribué à la déduplication des données mises à zéro.

De nombreuses baies concurrentes prétendent être capables de dédupliquer des bases de données en présumant qu'une base de données est copiée plusieurs fois. Il est également possible d'utiliser la déduplication NetApp, mais ONTAP offre une meilleure option : la technologie FlexClone de NetApp. Le résultat final est le même : plusieurs copies d'une base de données qui partagent la plupart des blocs physiques sous-jacents sont créées. L'utilisation de FlexClone est bien plus efficace que de prendre le temps de copier les fichiers de base de données, puis de les dédupliquer. Il s'agit en effet de la non-duplication plutôt que de la déduplication, car un doublon n'est jamais créé à la première place.

Efficacité et provisionnement fin

Les fonctions d'efficacité sont des formes de provisionnement fin. Par exemple, une LUN de 100 Go occupant un volume de 100 Go peut compresser à 50 Go. Aucune économie réelle n'est encore réalisée, car le volume est toujours de 100 Go. Le volume doit d'abord être réduit afin que l'espace économisé puisse être utilisé ailleurs sur le système. Si des modifications ultérieures de la LUN de 100 Go réduisent la taille des données compressibles, la LUN augmente et le volume pourrait se remplir.

Le provisionnement fin est fortement recommandé car il simplifie la gestion tout en améliorant la capacité exploitable avec les économies associées. La raison en est simple : les environnements de base de données

comportent souvent beaucoup d'espace vide, un grand nombre de volumes et de LUN, ainsi que des données compressibles. Le provisionnement fin entraîne la réservation d'espace sur le stockage pour les volumes et les LUN au cas où un jour ils se traduirait par une saturation de 100 % et contiendraient des données non compressibles à 100 %. Il est peu probable que cela se produise. Le provisionnement fin permet de récupérer et d'utiliser cet espace ailleurs. Il permet également de gérer la capacité en fonction du système de stockage lui-même, plutôt que de nombreux volumes et LUN plus petits.

Certains clients préfèrent utiliser le provisionnement lourd, soit pour des charges de travail spécifiques, soit généralement en fonction de pratiques opérationnelles et d'approvisionnement établies.

> Si un volume est configuré en mode « Thick provisioning », veillez à désactiver complètement toutes les fonctions d'efficacité de ce volume, y compris la décompression et la suppression de la déduplication à l'aide de la `sis undo` commande. Le volume ne doit pas apparaître en `volume efficiency show` sortie. Si c'est le cas, le volume est encore partiellement configuré pour les fonctions d'efficacité. Par conséquent, les garanties de remplacement fonctionnent différemment, ce qui augmente le risque que les dépassements de configuration entraînent un manque inattendu d'espace du volume, ce qui entraîne des erreurs d'E/S de la base de données.

Meilleures pratiques en matière d'efficacité

NetApp recommande ce qui suit :

AFF par défaut

Les volumes créés sur ONTAP et exécutés sur un système AFF 100 % Flash sont à allocation dynamique, avec l'activation de toutes les fonctionnalités d'efficacité à la volée. Bien que les bases de données ne bénéficient généralement pas de la déduplication et puissent inclure des données non compressibles, les paramètres par défaut conviennent néanmoins à la plupart des charges de travail. ONTAP est conçu pour traiter efficacement tous les types de données et de modèles d'E/S, qu'ils entraînent ou non des économies. Les valeurs par défaut ne doivent être modifiées que si les raisons sont parfaitement comprises et si un écart est bénéfique.

Recommandations générales

- Si les volumes et/ou les LUN ne sont pas à provisionnement fin, vous devez désactiver tous les paramètres d'efficacité car l'utilisation de ces fonctionnalités ne permet aucune économie et la combinaison du provisionnement lourd et de l'optimisation de l'espace peut provoquer des comportements inattendus, notamment des erreurs de manque d'espace.
- Si les données ne sont pas sujettes à des écrasements, par exemple avec des sauvegardes ou des journaux de transactions de base de données, vous pouvez atteindre une meilleure efficacité en activant TSSE avec une période de refroidissement faible.
- Certains fichiers peuvent contenir une quantité importante de données non compressibles, par exemple lorsque la compression est déjà activée au niveau de l'application, les fichiers sont cryptés. Si l'un de ces scénarios est vrai, envisagez de désactiver la compression pour permettre un fonctionnement plus efficace sur d'autres volumes contenant des données compressibles.
- N'utilisez pas la compression et la déduplication de 32 Ko pour les sauvegardes de bases de données. Voir la section Compression adaptative pour plus d'informations.

Compression des bases de données

SQL Server lui-même dispose également de fonctionnalités pour compresser et gérer efficacement les données. SQL Server prend actuellement en charge deux types de compression de données : la compression

de ligne et la compression de page.

La compression de ligne modifie le format de stockage des données. Par exemple, il change les entiers et les décimales au format de longueur variable au lieu de leur format natif de longueur fixe. Il remplace également les chaînes de caractères de longueur fixe par le format de longueur variable en éliminant les espaces vides. La compression de page implémente la compression de ligne et deux autres stratégies de compression (compression de préfixe et compression de dictionnaire). Vous trouverez plus de détails sur la compression de page dans "Mise en œuvre de la compression de page".

La compression des données est actuellement prise en charge dans les éditions entreprise, Développeur et évaluation de SQL Server 2008 et versions ultérieures. Bien que la compression puisse être effectuée par la base de données elle-même, elle est rarement observée dans un environnement SQL Server.

Voici les recommandations pour la gestion de l'espace pour les fichiers de données SQL Server

- Utiliser le provisionnement fin dans les environnements SQL Server pour améliorer l'utilisation de l'espace et réduire les besoins globaux en stockage lorsque la fonctionnalité de garantie d'espace est utilisée.
  - Utilisez le croissance automatique dans la plupart des configurations de déploiement courantes, car l'administrateur du stockage ne doit contrôler l'utilisation de l'espace dans l'agrégat.
- N'activez pas la déduplication sur les volumes FAS contenant des fichiers de données SQL Server, sauf si le volume contient plusieurs copies des mêmes données, telles que la restauration d'une base de données à partir de sauvegardes sur un seul volume.

Réclamations d'espace

La récupération d'espace peut être lancée régulièrement pour restaurer l'espace inutilisé d'une LUN. Avec SnapCenter, vous pouvez utiliser la commande PowerShell suivante pour démarrer la récupération d'espace.

```
Invoke-SdHostVolumeSpaceReclaim -Path drive_path
```

Si vous devez exécuter la récupération d'espace, ce processus doit être exécuté pendant les périodes de faible activité car il consomme initialement des cycles sur l'hôte.

Protection des données

Les stratégies de sauvegarde des bases de données doivent être basées sur des exigences métier identifiées, et non sur des capacités théoriques. En combinant la technologie Snapshot de ONTAP et en exploitant les API de Microsoft SQL Server, vous pouvez effectuer rapidement des sauvegardes cohérentes au niveau des applications, quelle que soit la taille des bases de données utilisateur. Pour une gestion des données plus avancée ou scale-out, NetApp propose SnapCenter.

SnapCenter

SnapCenter est le logiciel NetApp de protection des données pour les applications d'entreprise. Les bases de données SQL Server peuvent être protégées rapidement et facilement grâce au plug-in SnapCenter pour SQL Server et aux opérations du système d'exploitation gérées par le plug-in SnapCenter pour Microsoft Windows.

L'instance SQL Server peut être une instance de cluster d'installation autonome ou de basculement, ou elle peut être toujours sur le groupe de disponibilité. Le résultat est que depuis une fenêtre unique, les bases de données peuvent être protégées, clonées et restaurées à partir d'une copie principale ou secondaire.

SnapCenter peut gérer les bases de données SQL Server à la fois sur site, dans le cloud et dans des configurations hybrides.des copies de bases de données peuvent également être créées en quelques minutes sur l'hôte original ou alternatif à des fins de développement ou de reporting.

SQL Server nécessite également une coordination entre le système d'exploitation et le stockage pour s'assurer que les données correctes sont présentes dans les snapshots au moment de la création. Dans la plupart des cas, la seule méthode sûre pour ce faire est SnapCenter ou T-SQL. Les snapshots créés sans cette coordination supplémentaire peuvent ne pas être récupérables de manière fiable.

Pour plus d'informations sur le plug-in SQL Server pour SnapCenter, reportez-vous à la section "Tr-4714 : guide des meilleures pratiques pour SQL Server avec NetApp SnapCenter".

Protection de la base de données à l'aide de snapshots T-SQL

Dans SQL Server 2022, Microsoft a introduit des snapshots T-SQL qui permettent de réaliser des scripts et d'automatiser les opérations de sauvegarde. Au lieu d'effectuer des copies complètes, vous pouvez préparer la base de données pour les snapshots. Une fois la base de données prête pour la sauvegarde, vous pouvez utiliser les API REST de ONTAP pour créer des snapshots.

Voici un exemple de flux de travail de sauvegarde :

1. Figez une base de données à l'aide de la commande ALTER. La base de données est ainsi préparée pour un snapshot cohérent sur le stockage sous-jacent. Après le gel, vous pouvez dégeler la base de données et enregistrer le snapshot avec la commande BACKUP.
2. Réalisez des instantanés de plusieurs bases de données sur les volumes de stockage simultanément avec les nouvelles commandes de GROUPE DE SAUVEGARDE et de SERVEUR DE SAUVEGARDE.
3. Effectuer des sauvegardes COMPLÈTES ou des sauvegardes COMPLÈTES COPY_ONLY. Ces sauvegardes sont également enregistrées dans msdb.
4. Effectuez une restauration instantanée à l'aide de sauvegardes de journaux effectuées avec l'approche de streaming standard après la sauvegarde COMPLÈTE des snapshots. Les sauvegardes différentielles en continu sont également prises en charge si nécessaire.

Pour en savoir plus, voir "Documentation Microsoft à connaître sur les snapshots T-SQL".

> **NetApp recommande** d'utiliser SnapCenter pour créer des copies Snapshot. La méthode T-SQL décrite ci-dessus fonctionne également, mais SnapCenter offre une automatisation complète du processus de sauvegarde, de restauration et de clonage. Il effectue également une découverte pour s'assurer que les snapshots corrects sont créés. Aucune préconfiguration n'est requise.

Groupe de disponibilité SQL Server avec SnapCenter

SnapCenter prend en charge la sauvegarde de la base de données du groupe de disponibilité SQL Server configurée avec un cluster de basculement Windows.

Le plug-in SnapCenter pour Microsoft SQL Server doit être installé sur tous les nœuds du cluster de basculement de serveur Windows. Reportez-vous à "documentation"la section sur les prérequis et aux étapes de configuration des plug-ins SnapCenter.

SnapCenter découvre toutes les bases de données, instances et groupes de disponibilité des hôtes Windows et les ressources sont énumérées sur la page de ressources SnapCenter.

Protection des bases de données dans le groupe disponibilité permanente

Les bases de données du groupe de disponibilité peuvent être protégées de plusieurs façons.

- Sauvegarde au niveau de la base de données : sélectionnez la base de données de la page de ressources de la base de données, ajoutez la stratégie consistant en une sauvegarde complète/journal, planifiez la sauvegarde. SnapCenter effectue la sauvegarde quel que soit le rôle de la base de données, qu'il s'agisse d'une réplique primaire ou secondaire. La protection peut également être configurée en ajoutant des bases de données au groupe de ressources.
- Sauvegarde au niveau de l'instance : sélectionnez l'instance et toutes les bases de données exécutées sur l'instance sont protégées en fonction de la règle sélectionnée. Toutes les bases de données, y compris la base de données de disponibilité s'exécutant comme réplique primaire ou secondaire, sont sauvegardées à l'aide de SnapCenter. La protection peut également être configurée en ajoutant une instance au groupe de ressources.
- Sauvegarde au niveau du groupe de disponibilité : lors de la configuration de la règle, SnapCenter dispose d'une option avancée pour la sauvegarde au niveau du groupe de disponibilité. Le paramètre de groupe de disponibilité dans la stratégie permet aux utilisateurs de sélectionner la préférence de réplique pour la sauvegarde. Vous pouvez sélectionner une réplique principale, secondaire ou toutes. L'option par défaut est basée sur le jeu de répliques de sauvegarde dans la configuration du groupe de disponibilité SQL Server.

Le paramètre de groupe de disponibilité dans la stratégie SnapCenter s'applique uniquement si la sauvegarde au niveau du groupe de disponibilité est utilisée pour protéger les bases de données du groupe de disponibilité et ne s'applique pas à la sauvegarde au niveau de la base de données ou de l'instance.

NetApp recommande d'utiliser la sauvegarde au niveau de la disponibilité pour sauvegarder toutes les répliques exécutées sur le stockage NetApp ONTAP.

Configuration de la sauvegarde des journaux dans SnapCenter

Si le groupe de disponibilité est configuré sur une configuration autonome de SQL Server, un disque dédié doit être monté sur chaque nœud d'un cluster de basculement de serveur Windows. Le disque dédié doit être utilisé pour configurer le répertoire des journaux afin d'enregistrer les sauvegardes du journal des transactions.

Si le groupe de disponibilité est configuré sur le cluster de basculement SQL Server, le disque en cluster doit être créé sur l'instance de cluster de basculement SQL Server vers le répertoire de journalisation de l'hôte.

Restauration de la base de données dans la configuration du groupe de disponibilité avec SnapCenter

- SnapCenter fournit l'option de réamorçage pour restaurer automatiquement la base de données à partir du dernier snapshot disponible sur le réplica secondaire. L'opération de réamorçage restaure automatiquement et joint la sauvegarde de la base de données au groupe de disponibilité.
- Une autre façon de restaurer la base de données de réplica dans le groupe de disponibilité consiste à rompre le groupe de disponibilité et à effectuer la restauration complète et complète du journal. Utilisez SnapCenter pour restaurer la base de données en mode norecovery, puis utilisez SQL Server Management Studio ou T-SQL pour rejoindre la base de données au groupe de disponibilité.
- Pour restaurer un sous-ensemble de données uniquement, la fonctionnalité de clonage de SnapCenter peut être utilisée pour créer une copie de clone de la base de données. La copie de la base de données est créée en quelques minutes à l'aide de SnapCenter, puis exporte les données vers le réplica principal à l'aide des outils natifs du serveur SQL.

Pour connaître les bonnes pratiques de configuration de l'infrastructure de stockage de la base de données

afin de répondre aux exigences RTO et RPO, consultez "Tr-4714 meilleures pratiques pour Microsoft SQL Server avec NetApp SnapCenter"la section .

> SnapCenter ne prend pas en charge les groupes de disponibilité distribués et les groupes de disponibilité contenus.

Reprise après incident

Reprise après incident

Les bases de données d'entreprise et les infrastructures applicatives ont souvent besoin d'une réplication pour se protéger contre les catastrophes naturelles ou les perturbations imprévues, avec un temps d'interruption minimal.

La fonction de réplication de groupe de disponibilité en continu de SQL Server peut constituer une excellente option, et NetApp offre des options pour intégrer la protection des données à la disponibilité continue. Toutefois, dans certains cas, il peut être intéressant d'opter pour la technologie de réplication ONTAP. Il existe trois options de base.

SnapMirror

La technologie SnapMirror offre une solution d'entreprise rapide et flexible pour la réplication de données sur des réseaux LAN et WAN. La technologie SnapMirror transfère uniquement les blocs de données modifiés vers la destination après la création du miroir initial, ce qui réduit considérablement les besoins en bande passante réseau. Il peut être configuré en mode synchrone ou asynchrone.

Synchronisation active NetApp MetroCluster et SnapMirror

Pour de nombreux clients, la reprise après incident ne suffit pas à posséder une copie distante des données. Il est donc impératif de pouvoir les exploiter rapidement. NetApp propose deux technologies pour répondre à ce besoin : MetroCluster et SnapMirror Active Sync

MetroCluster fait référence à ONTAP dans une configuration matérielle qui inclut un stockage en miroir synchrone de faible niveau et de nombreuses fonctionnalités supplémentaires. Les solutions intégrées telles que MetroCluster simplifient les bases de données, les applications et les infrastructures de virtualisation complexes et évolutives. Elle remplace plusieurs produits et stratégies externes de protection des données par une seule baie de stockage centrale simple. Elle offre également des fonctionnalités intégrées de sauvegarde, de restauration, de reprise après incident et de haute disponibilité au sein d'un seul système de stockage en cluster.

La synchronisation active SnapMirror est basée sur SnapMirror synchrone. Avec MetroCluster, chaque contrôleur ONTAP est responsable de la réplication des données de son disque vers un emplacement distant. Avec la synchronisation active SnapMirror, deux systèmes ONTAP différents conservent des copies indépendantes de vos données LUN, mais fonctionnent ensemble pour présenter une seule instance de ce LUN. Du point de vue de l'hôte, il s'agit d'une entité LUN unique.

Comparaison SM-AS et MCC

Si les solutions SM-AS et MetroCluster sont similaires en termes de fonctionnalité globale, elles présentent d'importantes différences dans la mise en œuvre de la réplication avec un objectif de point de récupération de 0 et sa gestion. Les modes asynchrone et synchrone de SnapMirror peuvent également être utilisés dans le cadre d'un plan de reprise d'activité, mais ils ne sont pas conçus pour être utilisés en tant que technologies de réplication haute disponibilité.

- Une configuration MetroCluster ressemble davantage à un cluster intégré avec des nœuds distribués sur plusieurs sites. SM-AS se comporte comme deux clusters indépendants qui coopèrent pour fournir des LUN répliquées synchrones avec RPO=0 sélectionnés.
- Les données d'une configuration MetroCluster ne sont accessibles qu'à partir d'un site particulier à la fois. Une deuxième copie des données est présente sur le site opposé, mais les données sont passives. Il est impossible d'y accéder sans un basculement du système de stockage.
- La mise en miroir des systèmes MetroCluster et SM-AS effectue des opérations à différents niveaux. La mise en miroir MetroCluster s'effectue au niveau de la couche RAID. Les données de bas niveau sont stockées dans un format miroir à l'aide de SyncMirror. L'utilisation de la mise en miroir est pratiquement invisible au niveau des couches LUN, volume et protocole.
- En revanche, la mise en miroir SM-AS se produit au niveau de la couche de protocole. Les deux clusters sont globalement indépendants. Une fois les deux copies de données synchronisées, les deux clusters n'ont besoin que de mettre en miroir les écritures. Lorsqu'une écriture a lieu sur un cluster, elle est répliquée sur l'autre. L'écriture est uniquement validée par l'hôte lorsque l'écriture est terminée sur les deux sites. En dehors de ce comportement de fractionnement de protocole, les deux clusters sont des clusters ONTAP normaux.
- Le rôle principal de MetroCluster est la réplication à grande échelle. Vous pouvez répliquer une baie complète avec un objectif de point de récupération RPO=0 et un objectif de durée de restauration proche de zéro. Le processus de basculement est ainsi simplifié, car il n'y a qu'une seule « chose » à basculer et il offre une excellente évolutivité en termes de capacité et d'IOPS.
- L'une des principales utilisations de SM-AS est la réplication granulaire. Parfois, vous ne souhaitez pas répliquer toutes les données en tant qu'unité unique ou vous devez pouvoir basculer sélectivement sur certains workloads.
- Autre cas d'utilisation clé de la solution SM-as pour les opérations actives/actives : vous souhaitez que des copies de données entièrement exploitables soient disponibles sur deux clusters différents situés à deux emplacements différents avec des performances identiques et, si vous le souhaitez, vous n'avez pas besoin d'étendre le SAN sur plusieurs sites. Vos applications peuvent déjà s'exécuter sur les deux sites, ce qui réduit le RTO global pendant les opérations de basculement.

SnapMirror

Voici les recommandations pour SnapMirror pour SQL Server :

- Si SMB est utilisé, le SVM de destination doit être membre du même domaine Active Directory dont le SVM source est membre, de sorte que les listes de contrôle d'accès (ACL) stockées dans les fichiers NAS ne soient pas interrompues pendant la reprise après un incident.
- L'utilisation de noms de volume de destination identiques aux noms de volume source n'est pas requise, mais peut faciliter la gestion du processus de montage des volumes de destination dans la destination. Si SMB est utilisé, vous devez rendre l'espace de noms NAS de destination identique dans les chemins et la structure de répertoires vers l'espace de noms source.
- À des fins de cohérence, ne planifiez pas les mises à jour SnapMirror depuis les contrôleurs. Activez plutôt les mises à jour SnapMirror depuis SnapCenter pour mettre à jour SnapMirror une fois la sauvegarde complète ou la sauvegarde du journal terminée.
- Distribuez les volumes contenant des données SQL Server sur différents nœuds du cluster pour permettre à tous les nœuds de cluster de partager l'activité de réplication SnapMirror. Cette distribution optimise l'utilisation des ressources du nœud.
- Utiliser la réplication synchrone où la demande de restauration rapide des données est plus élevée et les solutions asynchrones pour plus de flexibilité dans le RPO.

Pour plus d'informations sur SnapMirror, reportez-vous à la section "Tr-4015 : Guide de configuration et des

meilleures pratiques de SnapMirror pour ONTAP 9".

MetroCluster

Architecture

Le déploiement de Microsoft SQL Server avec un environnement MetroCluster nécessite une explication de la conception physique d'un système MetroCluster.

MetroCluster met en miroir les données et la configuration de manière synchrone entre deux clusters ONTAP dans des emplacements distincts ou dans des domaines à défaillance. MetroCluster fournit aux applications du stockage disponible en continu en gérant automatiquement deux objectifs :

- Objectif de point de restauration (RPO) nul via une mise en miroir synchrone des données écrites sur le cluster.
- Objectif de délai de restauration (RTO) proche de zéro grâce à la mise en miroir de la configuration et à l'automatisation de l'accès aux données sur le second site.

MetroCluster simplifie la mise en miroir automatique des données et la configuration entre les deux clusters indépendants situés dans les deux sites. Le stockage étant provisionné dans un cluster, il est automatiquement mis en miroir sur le second cluster sur le second site. NetApp SyncMirror® fournit une copie complète de toutes les données avec un RPO nul. Cela signifie que les charges de travail d'un site peuvent basculer à tout moment sur le site opposé et continuer à transférer des données sans perte de données. MetroCluster gère le processus de basculement pour l'accès aux données provisionnées NAS et SAN sur le second site. La conception de la solution validée MetroCluster inclut le dimensionnement et la configuration qui permettent d'effectuer un basculement dans les délais impartis (en général moins de 120 secondes). Il en résulte un RPO proche de zéro et les applications peuvent continuer à accéder aux données sans produire de défaillances.MetroCluster est disponible selon plusieurs variations définies par le fabric de stockage interne.

MetroCluster est disponible dans 3 configurations différentes

- Paires HAUTE DISPONIBILITÉ avec connectivité IP
- Paires HAUTE DISPONIBILITÉ avec connectivité FC
- Contrôleur unique avec connectivité FC

> Le terme « connectivité » fait référence à la connexion au cluster utilisée pour la réplication entre sites. Il ne fait pas référence aux protocoles hôtes. Tous les protocoles côté hôte sont pris en charge comme d'habitude dans une configuration MetroCluster, quel que soit le type de connexion utilisé pour les communications entre clusters.

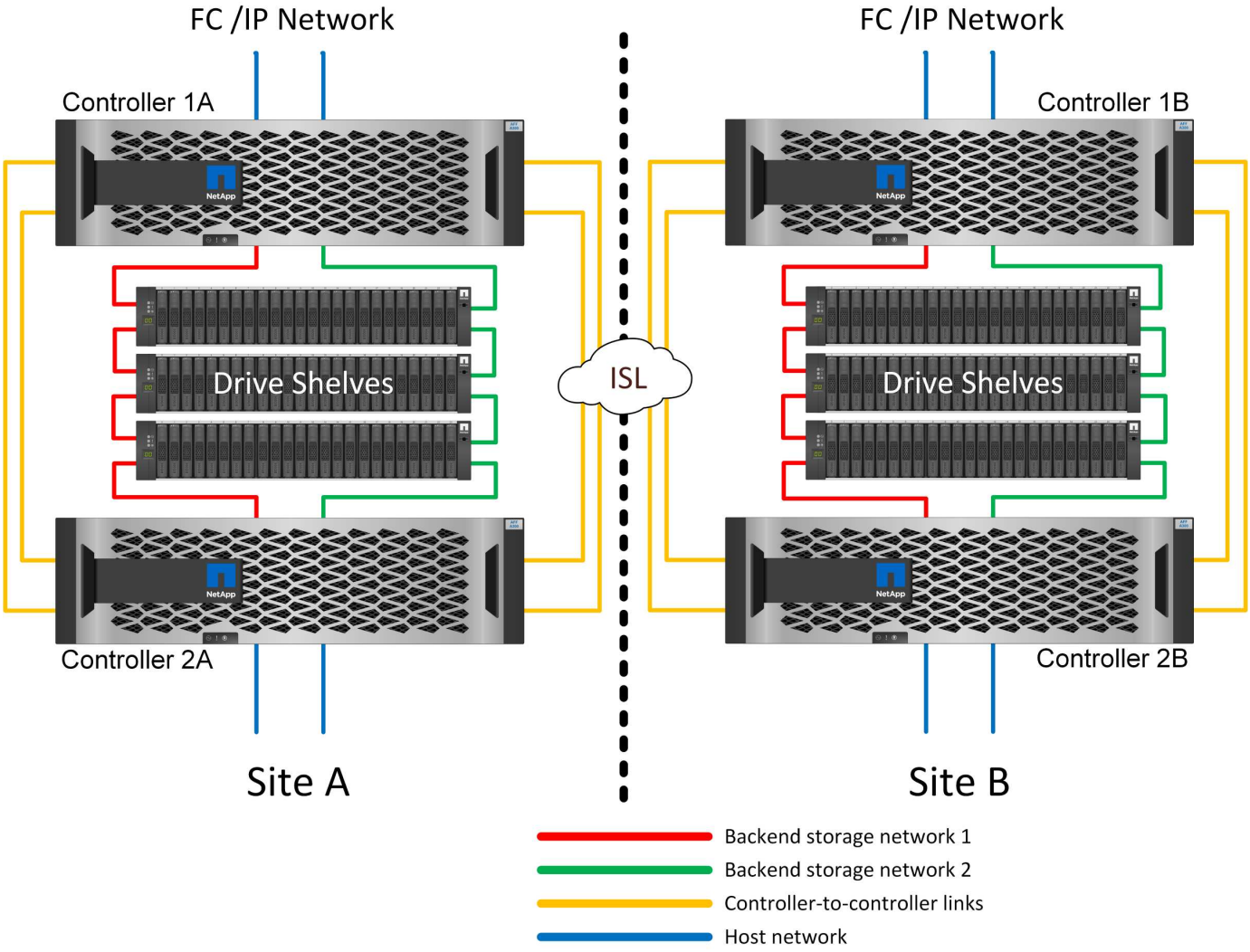
IP MetroCluster

La configuration IP MetroCluster à paire haute disponibilité utilise deux ou quatre nœuds par site. Cette option de configuration augmente la complexité et les coûts liés à l'option à deux nœuds, mais elle offre un avantage important : la redondance intrasite. Une simple panne de contrôleur ne nécessite pas l'accès aux données via le WAN. L'accès aux données reste local via l'autre contrôleur local.

La plupart des clients choisissent la connectivité IP, car les exigences d'infrastructure sont plus simples. Auparavant, la connectivité inter-sites à haut débit était généralement plus facile à provisionner avec des commutateurs FC et fibre noire. Cependant, les circuits IP à haut débit et à faible latence sont aujourd'hui plus facilement disponibles.

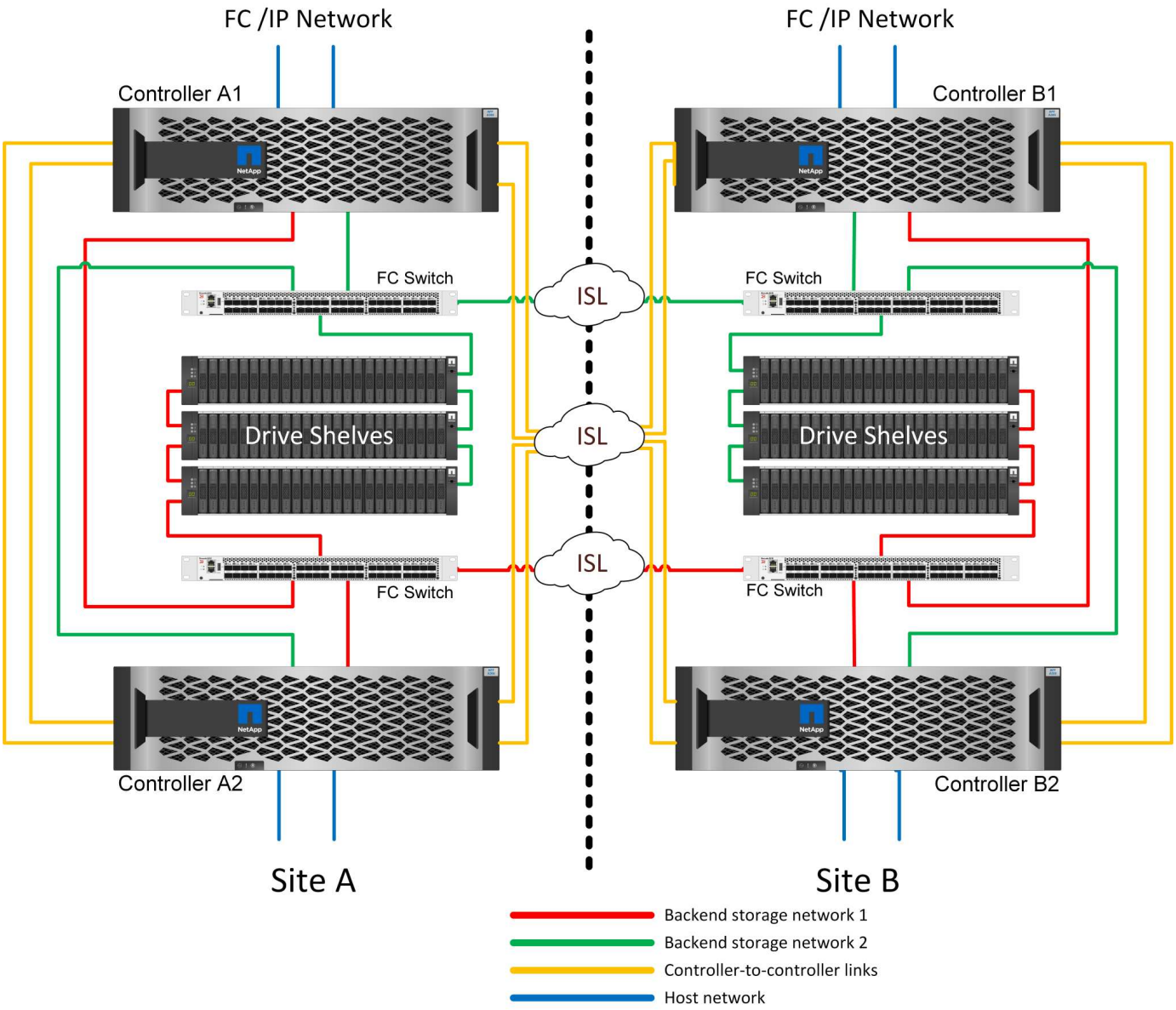
L'architecture est également plus simple, car les contrôleurs disposent des seules connexions entre les sites. Dans les MetroCluster FC, un contrôleur écrit directement sur les disques du site opposé et requiert ainsi des connexions SAN, des commutateurs et des ponts supplémentaires. En revanche, un contrôleur dans une configuration IP écrit sur les lecteurs opposés via le contrôleur.

Pour plus d'informations, consultez la documentation officielle de ONTAP et "Architecture et conception de la solution IP de MetroCluster".



MetroCluster FC à connexion SAN HA-pair

La configuration MetroCluster FC à paire haute disponibilité utilise deux ou quatre nœuds par site. Cette option de configuration augmente la complexité et les coûts liés à l'option à deux nœuds, mais elle offre un avantage important : la redondance intrasite. Une simple panne de contrôleur ne nécessite pas l'accès aux données via le WAN. L'accès aux données reste local via l'autre contrôleur local.



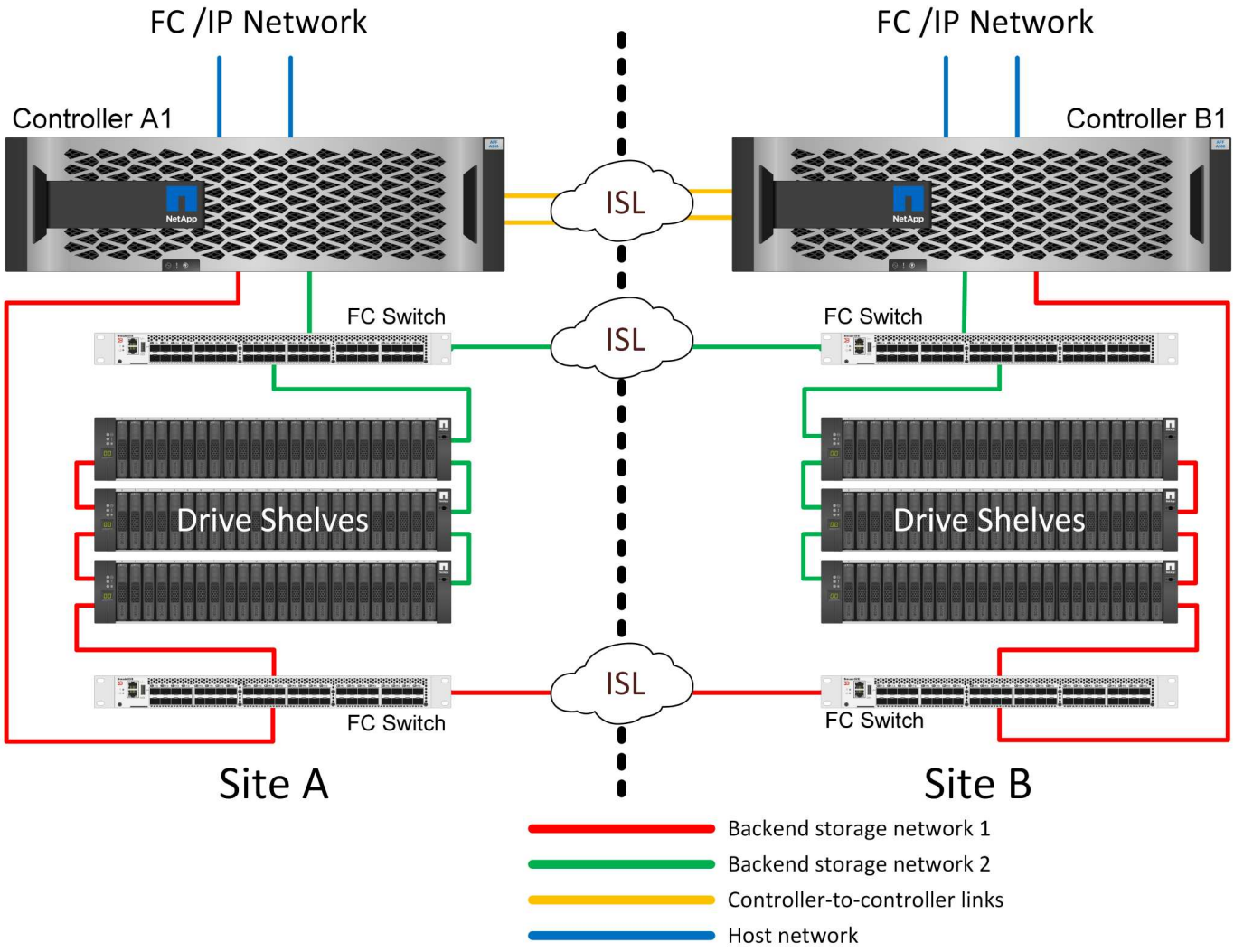
Certaines infrastructures multisites ne sont pas conçues pour les opérations en mode actif-actif. Elles sont plutôt utilisées comme site principal et site de reprise après incident. Dans ce cas, il est généralement préférable d'utiliser une option MetroCluster à paire HA pour les raisons suivantes :

- Bien qu'un cluster MetroCluster à deux nœuds soit un système haute disponibilité, toute panne inattendue d'un contrôleur ou une maintenance planifiée implique que les services de données soient en ligne sur le site opposé. Si la connectivité réseau entre les sites ne prend pas en charge la bande passante requise, les performances sont affectées. La seule option serait également de basculer les différents systèmes d'exploitation hôtes et les services associés vers le site secondaire. Le cluster MetroCluster de paire haute disponibilité élimine ce problème, car la perte d'un contrôleur simplifie le basculement au sein du même site.
- Certaines topologies réseau ne sont pas conçues pour l'accès intersite, mais utilisent des sous-réseaux différents ou des SAN FC isolés. Dans ce cas, le cluster MetroCluster à deux nœuds ne fonctionne plus comme un système haute disponibilité, car le contrôleur secondaire ne peut plus transmettre de données aux serveurs sur le site opposé. L'option MetroCluster de paire haute disponibilité est nécessaire pour assurer une redondance complète.
- Si une infrastructure à deux sites est considérée comme une seule infrastructure extrêmement disponible, la configuration MetroCluster à deux nœuds est adaptée. Toutefois, si le système doit fonctionner pendant

une période prolongée après une panne sur le site, une paire haute disponibilité est recommandée, car la haute disponibilité continue d'être disponible sur un seul site.

MetroCluster FC à deux nœuds avec connexion SAN

La configuration MetroCluster à deux nœuds n'utilise qu'un nœud par site. Cette conception est plus simple que l'option de paire haute disponibilité, car le nombre de composants à configurer et à gérer est inférieur. Elle a également réduit les besoins en infrastructure en termes de câblage et de commutation FC. Enfin, il réduit les coûts.



L'impact évident de cette conception est que la défaillance du contrôleur sur un seul site signifie que les données sont disponibles depuis le site opposé. Cette restriction n'est pas nécessairement un problème. De nombreuses entreprises disposent d'opérations de data Center multisites avec des réseaux étendus, ultra-rapides et à faible latence qui fonctionnent essentiellement comme une infrastructure unique. Dans ce cas, la version à deux nœuds de MetroCluster est la configuration préférée. Plusieurs fournisseurs de services utilisent actuellement des systèmes à deux nœuds de plusieurs pétaoctets.

Fonctions de résilience MetroCluster

Une solution MetroCluster ne présente aucun point de défaillance unique :

- Chaque contrôleur dispose de deux chemins d'accès indépendants aux tiroirs disques sur le site local.

- Chaque contrôleur dispose de deux chemins d'accès indépendants aux tiroirs disques du site distant.
- Chaque contrôleur dispose de deux chemins d'accès indépendants aux contrôleurs sur le site opposé.
- Dans la configuration HA-pair, chaque contrôleur dispose de deux chemins vers son partenaire local.

En résumé, n'importe quel composant de la configuration peut être supprimé sans compromettre la capacité de MetroCluster à transmettre des données. La seule différence en termes de résilience entre les deux options est que la version à paire haute disponibilité reste un système de stockage haute disponibilité global après une panne de site.

SyncMirror

La protection pour SQL Server avec MetroCluster repose sur SyncMirror, qui offre une technologie de mise en miroir synchrone à évolutivité horizontale et aux performances maximales.

Protection des données avec SyncMirror

Au niveau le plus simple, la réplication synchrone implique que toute modification doit être apportée des deux côtés du stockage en miroir avant d'être reconnue. Par exemple, si une base de données écrit un journal ou si un invité VMware est en cours de correction, une écriture ne doit jamais être perdue. Au niveau du protocole, le système de stockage ne doit pas accuser réception de l'écriture tant qu'il n'a pas été validé sur un support non volatile des deux sites. Ce n'est qu'à cette condition qu'il est possible de continuer sans risque de perte de données.

L'utilisation d'une technologie de réplication synchrone est la première étape de la conception et de la gestion d'une solution de réplication synchrone. Il est important de comprendre ce qui pourrait se passer lors de divers scénarios de défaillance planifiés ou non. Les solutions de réplication synchrone offrent toutes des fonctionnalités différentes. Si vous avez besoin d'une solution avec un objectif de point de récupération de zéro, c'est-à-dire sans perte de données, tous les scénarios de défaillance doivent être pris en compte. En particulier, quel est le résultat escompté lorsque la réplication est impossible en raison d'une perte de connectivité entre les sites ?

Disponibilité des données SyncMirror

La réplication MetroCluster repose sur la technologie NetApp SyncMirror, conçue pour basculer efficacement en mode synchrone et en sortir. Cette fonctionnalité répond aux exigences des clients qui demandent une réplication synchrone, mais qui ont également besoin d'une haute disponibilité pour leurs services de données. Par exemple, si la connectivité à un site distant est coupée, il est généralement préférable que le système de stockage continue de fonctionner dans un état non répliqué.

De nombreuses solutions de réplication synchrone ne peuvent fonctionner qu'en mode synchrone. Ce type de réplication « tout ou rien » est parfois appelé mode domino. Ces systèmes de stockage cessent d'accéder aux données au lieu d'interrompre la synchronisation des copies locales et distantes des données. Si la réplication est forcée, la resynchronisation peut prendre beaucoup de temps et laisser un client exposé à des pertes de données complètes pendant la période de rétablissement de la mise en miroir.

Non seulement SyncMirror peut basculer en mode synchrone sans interruption si le site distant est inaccessible, mais il peut également rapidement resynchroniser vers un état RPO = 0 une fois la connectivité restaurée. La copie obsolète des données sur le site distant peut également être conservée dans un état utilisable lors de la resynchronisation, garantissant la présence à tout moment de copies locales et distantes des données.

Si le mode domino est requis, NetApp propose SnapMirror synchrone (SM-S). Des options au niveau de l'application existent également, comme Oracle DataGuard ou SQL Server Always On Availability Groups. La mise en miroir des disques au niveau du système d'exploitation peut être optionnelle. Pour plus d'informations

et d'options, consultez votre équipe de compte NetApp ou partenaire.

SQL Server avec MetroCluster

MetroCluster est une option de protection des bases de données SQL Server avec un RPO nul. MetroCluster est une technologie de réplication simple, haute performance avec un objectif de point de récupération de 0, qui vous permet de répliquer facilement toute une infrastructure sur plusieurs sites.

SQL Server peut évoluer jusqu'à des milliers de bases de données sur un seul système MetroCluster. Il peut y avoir des instances autonomes SQL Server ou des instances de cluster de basculement ; le système MetroCluster n'ajoute pas nécessairement ou ne modifie pas les meilleures pratiques de gestion d'une base de données.

Une explication complète de MetroCluster dépasse le champ d'application de ce document, mais les principes sont simples. MetroCluster peut fournir une solution de réplication RPO=0 avec basculement rapide. La base que vous exploitez dépend de vos besoins.

Par exemple, une procédure de reprise sur incident rapide de base après une perte soudaine du site peut utiliser les étapes de base suivantes :

- Force le basculement MetroCluster
- Découverte de LUN FC/iSCSI (SAN uniquement)
- Montez les systèmes de fichiers
- Démarrez SQL Services

Cette approche doit avant tout se passer d'un système d'exploitation en cours d'exécution sur le site distant. Il doit être préconfiguré avec la configuration de SQL Server et doit être mis à jour avec une version de build équivalente. Les bases de données système SQL Server peuvent également être mises en miroir sur le site distant et montées en cas d'incident.

Si les volumes, les systèmes de fichiers et les datastores hébergeant des bases de données virtualisées ne sont pas utilisés sur le site de reprise d'activité avant le basculement, il n'est pas nécessaire de les définir `dr-force-nvfail` sur les volumes associés.

Synchronisation active SnapMirror

Présentation

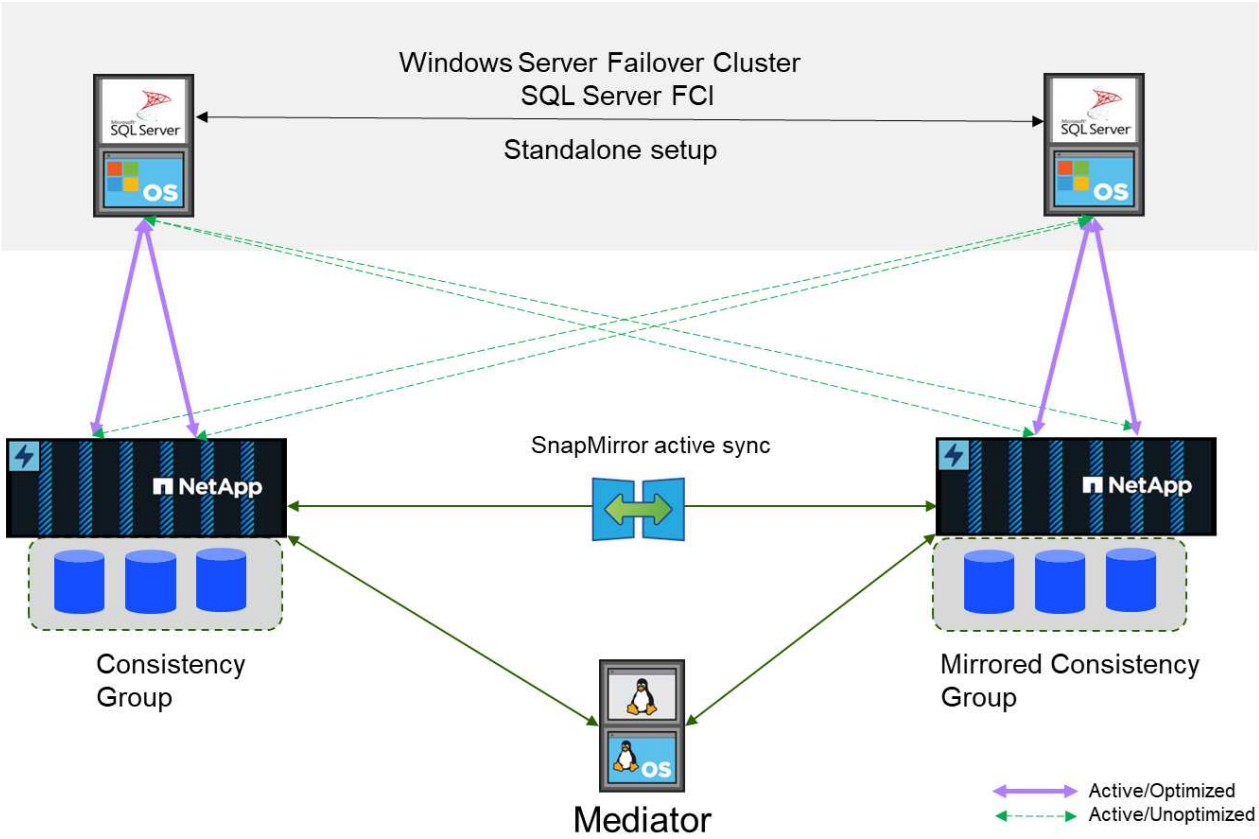
La synchronisation active SnapMirror permet à chaque base de données et application SQL Server de continuer les opérations pendant les interruptions du stockage et du réseau, grâce à un basculement transparent du stockage, sans intervention manuelle.

À partir de ONTAP 9.15.1, la synchronisation active SnapMirror prend en charge l'architecture active/active symétrique en plus de la configuration asymétrique existante. La fonctionnalité actif-actif symétrique offre une réplication bidirectionnelle synchrone pour la continuité de l'activité et la reprise après incident. Il vous aide à protéger l'accès aux données pour les workloads SAN stratégiques avec un accès simultané en lecture et en écriture aux données dans plusieurs domaines à défaillance. Vous bénéficiez ainsi d'une continuité de l'activité et d'une réduction des temps d'indisponibilité en cas d'incident ou de panne système.

Les hôtes SQL Server accèdent au stockage via des LUN Fibre Channel (FC) ou iSCSI. Réplication entre chaque cluster hébergeant une copie des données répliquées. Étant donné que cette fonctionnalité est une

réplication au niveau du stockage, les instances SQL Server exécutées sur des instances d'hôte autonome ou de cluster de basculement peuvent effectuer des opérations de lecture/écriture sur l'un ou l'autre des clusters. Pour les étapes de planification et de configuration, reportez-vous à la section "Documentation ONTAP sur la synchronisation active SnapMirror" .

Architecture de la synchronisation active SnapMirror avec actif-actif symétrique



Réplication synchrone

En fonctionnement normal, chaque copie correspond à une réplique synchrone RPO=0 à tout moment, à une exception près. Si les données ne peuvent pas être répliquées, ONTAP exige de répliquer les données et de reprendre le traitement des E/S sur un site pendant que les LUN de l'autre site sont mises hors ligne.

Matériel de stockage

Contrairement à d'autres solutions de reprise après incident du stockage, la synchronisation active SnapMirror offre une flexibilité asymétrique de la plateforme. Le matériel de chaque site n'a pas besoin d'être identique. Cette fonctionnalité vous permet d'ajuster la taille du matériel utilisé pour prendre en charge la synchronisation active SnapMirror. Le système de stockage distant peut être identique au site principal s'il doit prendre en charge une charge de travail de production complète, mais si un incident entraîne une réduction des E/S, un système plus petit sur le site distant peut être plus économique.

Médiateur ONTAP

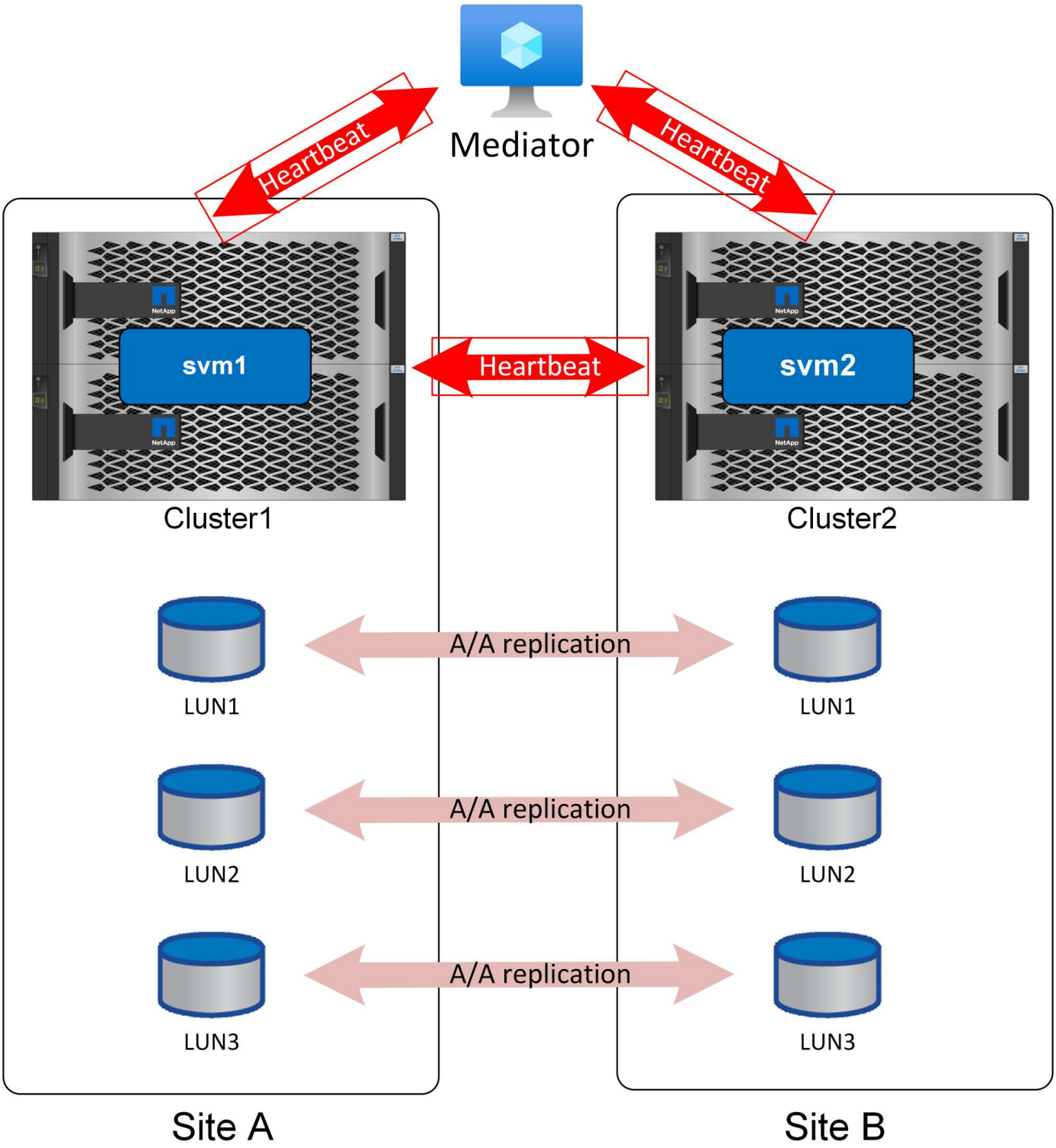
Le médiateur ONTAP est une application logicielle téléchargée depuis la prise en charge de NetApp et généralement déployée sur une petite machine virtuelle. Le médiateur ONTAP n'est pas un tiebreaker. Il s'agit d'un canal de communication alternatif pour les deux clusters qui participent à la réplication SnapMirror active Sync. Les opérations automatisées sont dirigées par ONTAP sur la base des réponses reçues du partenaire

via des relations directes et via le médiateur.

ONTAP Médiateur

Le Mediator est requis pour automatiser le basculement en toute sécurité. Dans l'idéal, elle serait placée sur un site tiers indépendant, mais elle peut toujours fonctionner pour la plupart des besoins si elle est en colocation avec l'un des clusters participant à la réplication.

Le médiateur n'est pas vraiment un casse-barre, bien que c'est effectivement la fonction qu'il fournit. Il ne prend aucune action ; il fournit plutôt un canal de communication alternatif pour la communication cluster à cluster.



Le principal défi lié au basculement automatisé est le problème des réseaux partagés, qui se pose en cas de perte de connectivité entre les deux sites. Que doit-on faire ? Vous ne voulez pas que deux sites différents se désignent comme les copies restantes des données, mais comment un seul site peut-il faire la différence entre la perte réelle du site opposé et l'incapacité à communiquer avec le site opposé ?

C'est là que le médiateur entre dans la photo. S'il est placé sur un troisième site, et chaque site a une connexion réseau distincte à ce site, alors vous avez un chemin supplémentaire pour chaque site pour valider l'état de santé de l'autre. Examinez à nouveau l'image ci-dessus et examinez les scénarios suivants.

- Que se passe-t-il si le médiateur échoue ou est inaccessible à partir d'un ou des deux sites ?
  - Les deux clusters peuvent toujours communiquer entre eux sur le même lien que celui utilisé pour les services de réplication.
  - Les données restent protégées avec un objectif de point de récupération de 0
- Que se passe-t-il si le site A tombe en panne ?
  - Le site B verra les deux canaux de communication tomber en panne.
  - Le site B prendra le contrôle des services de données, mais sans mise en miroir RPO=0
- Que se passe-t-il si le site B tombe en panne ?
  - Le site A verra les deux canaux de communication tomber en panne.
  - Le site A prend le relais des services de données, mais sans mise en miroir avec un objectif de point de récupération de 0

Il y a un autre scénario à prendre en compte : la perte du lien de réplication des données. En cas de perte de la liaison de réplication entre les sites, la mise en miroir avec un objectif de point de récupération de 0 sera évidemment impossible. Que devrait-on alors se passer ?

Ceci est contrôlé par le statut du site préféré. Dans une relation SM-AS, l'un des sites est secondaire à l'autre. Cela n'a aucun effet sur les opérations normales, et tout accès aux données est symétrique. Toutefois, si la réplication est interrompue, le nœud devra être rompu pour reprendre les opérations. Par conséquent, le site privilégié continuera les opérations sans mise en miroir et le site secondaire arrêtera le traitement des E/S jusqu'à ce que la communication de réplication soit restaurée.

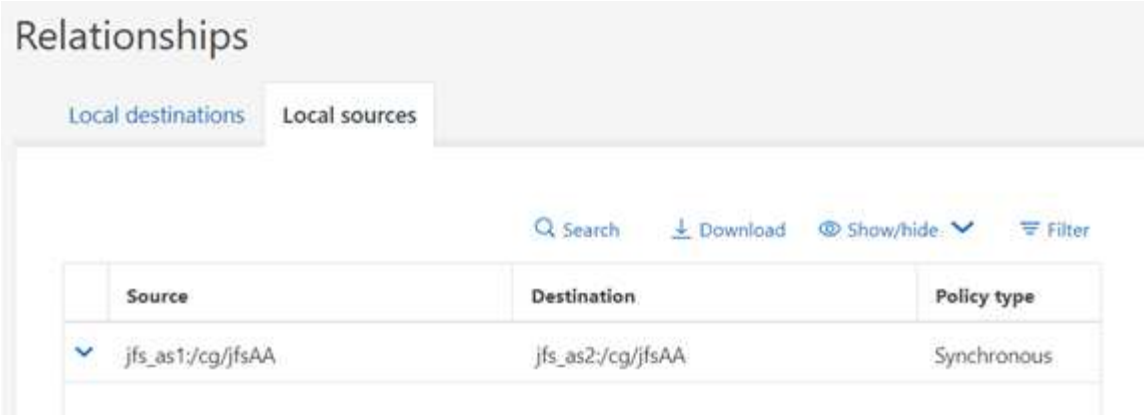
Site préféré

Le comportement de la synchronisation active SnapMirror est symétrique, avec une exception importante : la configuration du site préféré.

La synchronisation active SnapMirror considère un site comme la « source » et l'autre comme la « destination ». Cela implique une relation de réplication unidirectionnelle, mais cela ne s'applique pas au comportement d'E/S. La réplication est bidirectionnelle et symétrique. Les temps de réponse d'E/S sont identiques de part et d'autre du miroir.

La `source` désignation est le contrôle du site préféré. En cas de perte du lien de réplication, les chemins de LUN sur la copie source continueront à transmettre des données tandis que les chemins de LUN sur la copie de destination deviendront indisponibles jusqu'à ce que la réplication soit rétablie et que SnapMirror repasse à l'état synchrone. Les chemins reprennent alors le service des données.

La configuration source/destination peut être affichée via SystemManager :



Ou sur l'interface de ligne de commande :

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                      Destination Path: jfs_as2:/cg/jfsAA
                    Relationship Type: XDP
               Relationship Group Type: consistencygroup
                   SnapMirror Schedule: -
                SnapMirror Policy Type: automated-failover-duplex
                     SnapMirror Policy: AutomatedFailOverDuplex
                           Tries Limit: -
                     Throttle (KB/sec): -
                          Mirror State: Snapmirrored
                   Relationship Status: InSync
```

La clé est que la source est le SVM sur le cluster1. Comme mentionné ci-dessus, les termes « source » et « destination » ne décrivent pas le flux des données répliquées. Les deux sites peuvent traiter une écriture et la répliquer sur le site opposé. En effet, les deux grappes sont des sources et des destinations. La désignation d'un cluster comme source contrôle simplement le cluster qui survit en tant que système de stockage en lecture/écriture en cas de perte du lien de réplication.

Topologie réseau

Accès uniforme

Un réseau d'accès uniforme signifie que les hôtes peuvent accéder aux chemins sur les deux sites (ou domaines de défaillance au sein du même site).

L'une des caractéristiques importantes de SM-AS est la capacité de configurer les systèmes de stockage pour savoir où se trouvent les hôtes. Lorsque vous mappez les LUN sur un hôte donné, vous pouvez indiquer si elles sont proximales ou non à un système de stockage donné.

Paramètres de proximité

La proximité fait référence à une configuration par cluster qui indique qu'un WWN d'hôte ou un ID d'initiateur iSCSI appartient à un hôte local. Il s'agit d'une deuxième étape facultative de configuration de l'accès aux

LUN.

La première étape correspond à la configuration habituelle du groupe initiateur. Chaque LUN doit être mappée sur un groupe initiateur qui contient les ID WWN/iSCSI des hôtes devant accéder à cette LUN. Cela contrôle quel hôte a *accès* à un LUN.

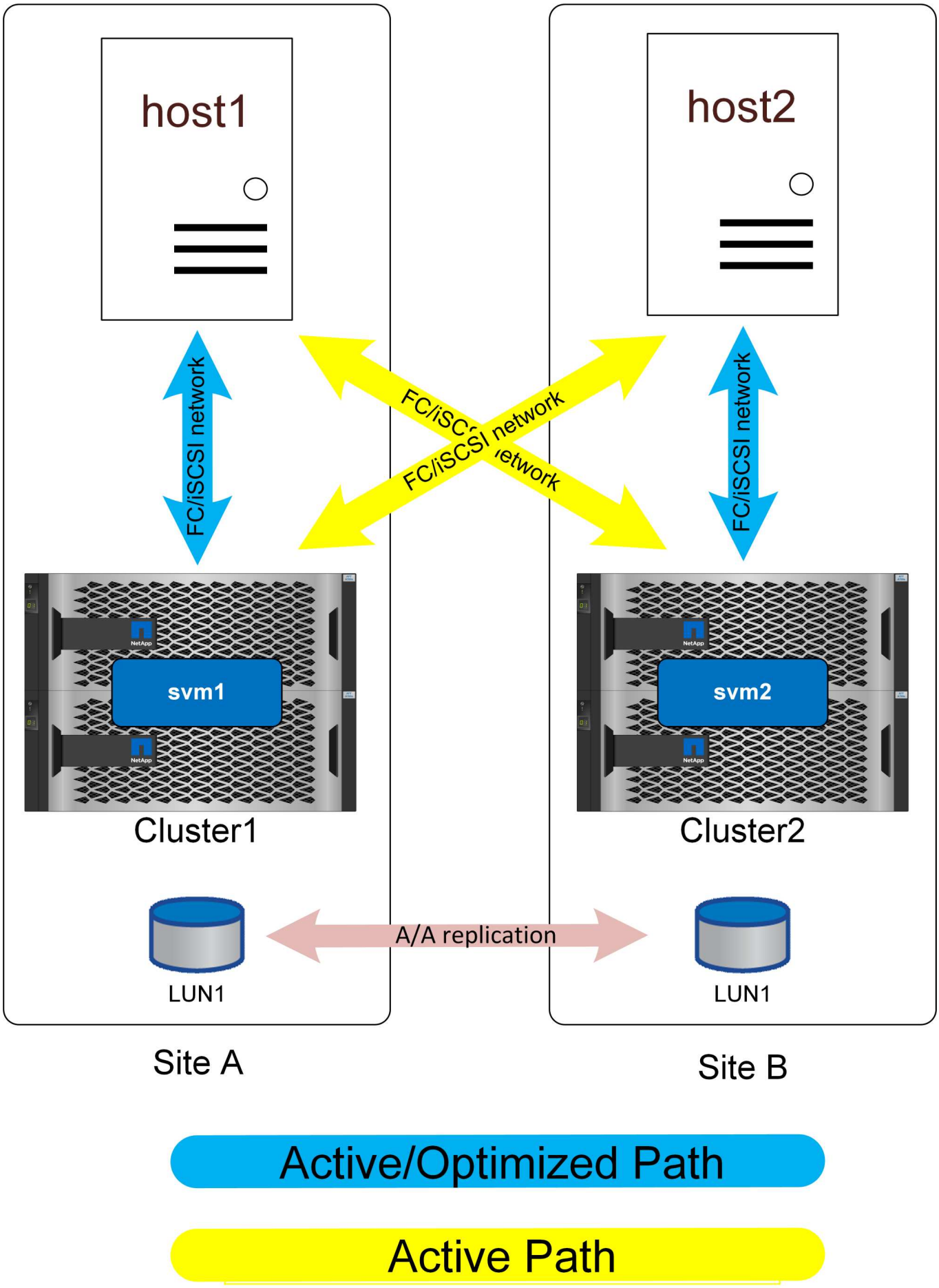
La deuxième étape facultative consiste à configurer la proximité de l'hôte. Cela ne contrôle pas l'accès, il contrôle *Priority*.

Par exemple, un hôte du site A peut être configuré pour accéder à une LUN protégée par la synchronisation active SnapMirror. Le SAN étant étendu entre les sites, les chemins d'accès sont disponibles pour cette LUN via le stockage sur le site A ou le stockage sur le site B.

Sans paramètres de proximité, cet hôte utilisera les deux systèmes de stockage de la même manière, car les deux systèmes de stockage annonceront des chemins actifs/optimisés. Si la latence SAN et/ou la bande passante entre les sites est limitée, il se peut que cela ne soit pas désirable, et vous pouvez vous assurer que, pendant le fonctionnement normal, chaque hôte utilise de préférence des chemins vers le système de stockage local. Cette configuration s'effectue en ajoutant l'ID WWN/iSCSI de l'hôte au cluster local en tant qu'hôte proximal. Cette opération peut être effectuée à partir de l'interface de ligne de commande ou de SystemManager.

AFF

Avec un système AFF, les chemins apparaissent comme indiqué ci-dessous lorsque la proximité de l'hôte a été configurée.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
svm2
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

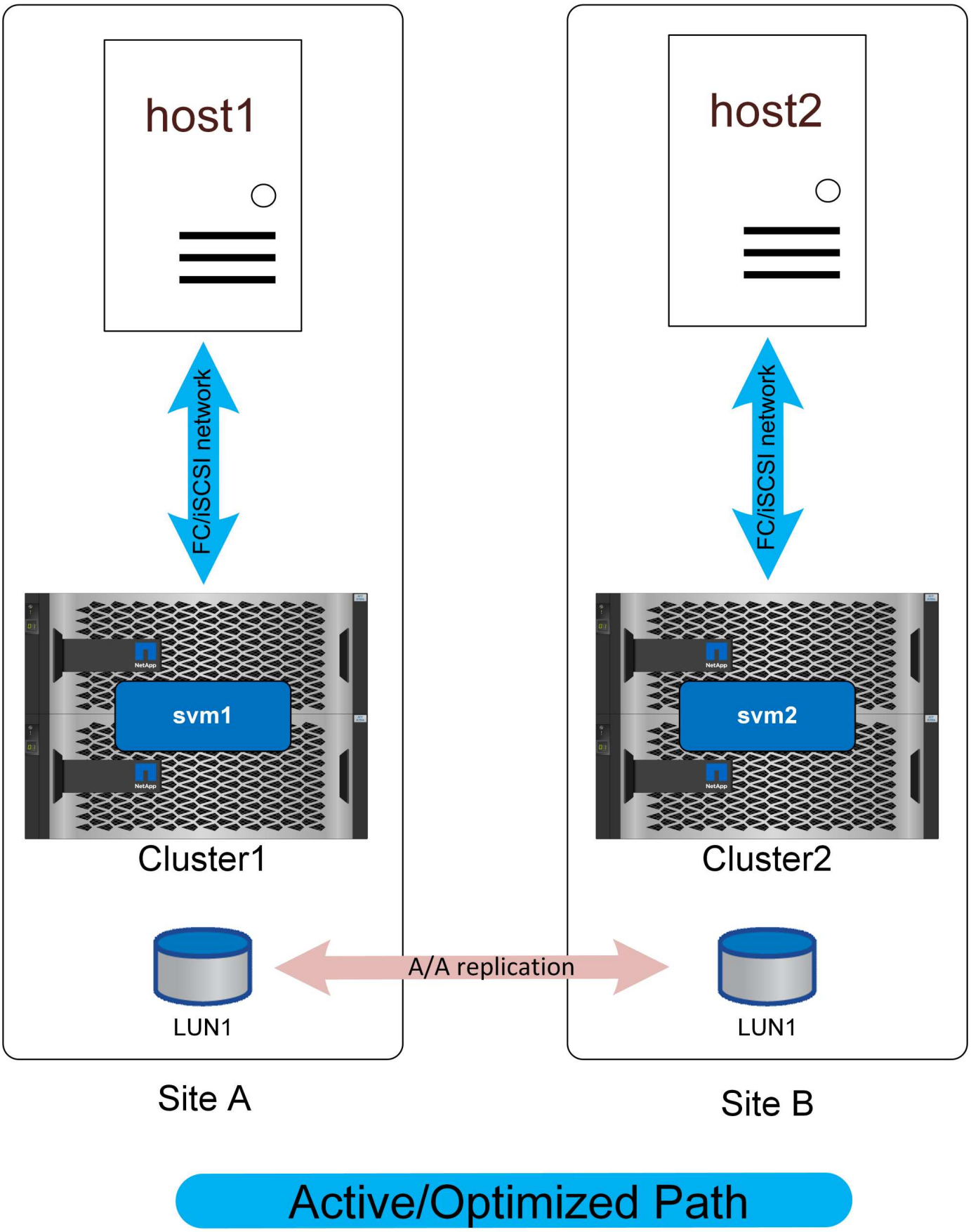
En fonctionnement normal, toutes les E/S sont des E/S locales. Les opérations de lecture et d'écriture sont gérées à partir de la baie de stockage locale. Bien entendu, les E/S en écriture devront également être répliquées par le contrôleur local sur le système distant avant d'être acquittées, mais toutes les E/S en lecture seront gérées localement et ne subiront pas de latence supplémentaire en traversant la liaison SAN entre les sites.

Le seul moment où les chemins non optimisés seront utilisés est la perte de tous les chemins actifs/optimisés. Par exemple, si l'ensemble de la baie sur le site A est hors tension, les hôtes du site A peuvent toujours accéder aux chemins d'accès à la baie sur le site B et donc rester opérationnels, même s'ils connaissent une latence plus élevée.

Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

Accès non uniforme

La mise en réseau à accès non uniforme signifie que chaque hôte n'a accès qu'aux ports du système de stockage local. Le SAN n'est pas étendu sur les sites (ou les domaines de défaillance au sein du même site).



Le principal avantage de cette approche est la simplicité du SAN : vous n'avez plus besoin d'étendre un SAN sur le réseau. Certains clients ne disposent pas d'une connectivité à faible latence suffisante entre les sites, ou

n’ont pas l’infrastructure nécessaire pour acheminer le trafic SAN FC sur un réseau intersite.

L’inconvénient de l’accès non uniforme est que certains scénarios de défaillance, notamment la perte du lien de réplication, entraînent la perte de l’accès au stockage par certains hôtes. En cas de perte de la connectivité du stockage local, les applications qui s’exécutent en tant qu’instances uniques, telles qu’une base de données non en cluster et qui ne s’exécute intrinsèquement que sur un hôte unique sur un montage donné, échouent. Les données seraient toujours protégées, mais le serveur de base de données n’aurait plus accès. Il doit être redémarré sur un site distant, de préférence par le biais d’un processus automatisé. Par exemple, VMware HA peut détecter une situation de tous les chemins d’accès sur un serveur et redémarrer une machine virtuelle sur un autre serveur sur lequel les chemins d’accès sont disponibles.

En revanche, une application en cluster telle qu’Oracle RAC peut fournir un service qui est disponible simultanément sur deux sites différents. La perte d’un site ne signifie pas la perte du service d’application dans son ensemble. Les instances restent disponibles et s’exécutent sur le site survivant.

Dans de nombreux cas, la surcharge liée à la latence supplémentaire qu’une application accède au système de stockage via une liaison site à site ne serait pas acceptable. Cela signifie que l’amélioration de la disponibilité des réseaux uniformes est minime, car la perte de stockage sur un site entraînerait la nécessité de fermer les services sur ce site défaillant.

> Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

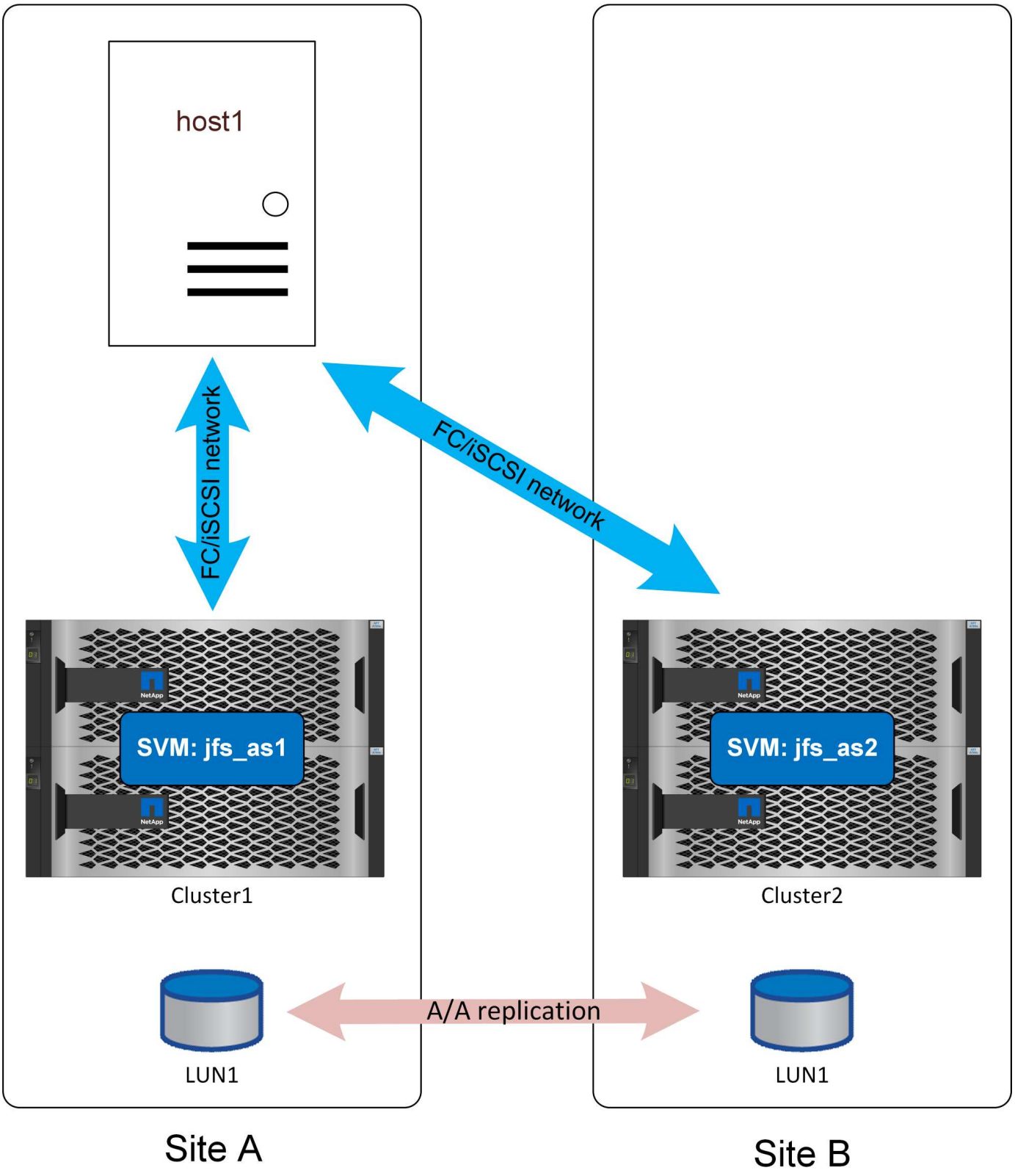
Présentation

SQL Server peut être configuré pour fonctionner avec la synchronisation active SnapMirror de plusieurs façons. La bonne réponse dépend de la connectivité réseau disponible, des exigences de RPO et de la disponibilité.

Instance autonome de SQL Server

Les meilleures pratiques en matière de mise en page des fichiers et de configuration des serveurs sont les mêmes que celles recommandées dans "SQL Server sur ONTAP" la documentation.

Avec une configuration autonome, SQL Server ne peut être exécuté que sur un site. L’accès serait probablement "uniforme"utilisé.



Avec un accès uniforme, une panne de stockage sur l'un ou l'autre site n'interromprait pas les opérations de la base de données. Une défaillance complète du site incluant le serveur de base de données entraînerait, bien sûr, une panne.

Certains clients peuvent configurer un système d'exploitation s'exécutant sur le site distant avec une configuration SQL Server préconfigurée, mise à jour avec une version de build équivalente à celle de l'instance

de production. Le basculement nécessite l'activation de cette instance autonome de SQL Server sur le site secondaire, la découverte des LUN et le démarrage de la base de données. Le processus complet peut être automatisé avec l'applet de commande Windows PowerShell, car aucune opération n'est requise côté stockage.

"Non uniforme" l'accès peut également être utilisé, mais il en résulte une panne de la base de données si le système de stockage sur lequel était situé le serveur de base de données avait échoué car la base de données ne disposait pas de chemins d'accès au stockage. Cela peut toujours être acceptable dans certains cas. La synchronisation active SnapMirror offre toujours une protection des données avec un objectif de point de récupération de 0. En cas de défaillance du site, la copie restante est active et prête à reprendre les opérations en suivant la même procédure utilisée avec un accès uniforme que celle décrite ci-dessus.

Un processus de basculement simple et automatisé peut être configuré plus facilement grâce à l'utilisation d'un hôte virtualisé. Par exemple, si les fichiers de données SQL Server sont répliqués de manière synchrone sur le stockage secondaire avec un VMDK de démarrage, l'environnement complet peut être activé sur l'autre site en cas d'incident. Un administrateur peut activer manuellement l'hôte sur le site survivant ou automatiser le processus via un service tel que VMware HA.

Instance de cluster de basculement SQL Server

Les instances de basculement SQL Server peuvent également être hébergées sur un cluster de basculement Windows s'exécutant sur un serveur physique ou virtuel en tant que système d'exploitation invité. Cette architecture multi-hôtes fournit l'instance SQL Server et la résilience du stockage. Ce déploiement est utile dans les environnements très exigeants qui recherchent des processus de basculement robustes tout en maintenant des performances améliorées. Dans une configuration de cluster de basculement, lorsqu'un hôte ou un stockage primaire est affecté, SQL Services effectue un basculement vers l'hôte secondaire et, dans le même temps, le stockage secondaire est disponible pour transmettre les E/S. Aucun script d'automatisation ni aucune intervention de l'administrateur n'est nécessaire.

Scénarios d'échec

La planification d'une architecture complète d'applications de synchronisation active SnapMirror nécessite de comprendre comment les SM-AS répondront dans divers scénarios de basculement planifiés et non planifiés.

Pour les exemples suivants, supposons que le site A est configuré comme le site préféré.

Perte de la connectivité de réplication

Si la réplication SM-AS est interrompue, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A (site préféré)

Le résultat de l'échec de la liaison de réplication sur le site préféré sera une pause d'environ 15 secondes dans le traitement des E/S d'écriture, car ONTAP relance les opérations d'écriture répliquées avant de déterminer que la liaison de réplication est véritablement inaccessible. Au bout de 15 secondes, le site A du système reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Panne du système de stockage

Le résultat d'une défaillance du système de stockage est presque identique au résultat de la perte du lien de réplication. Le site survivant devrait subir une pause d'E/S d'environ 15 seconde. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Perte du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ». En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

La raison pour laquelle un médiateur est nécessaire pour un basculement automatisé sécurisé est parce qu'il serait autrement impossible à un cluster de stockage de déterminer si la perte de la communication bidirectionnelle était le résultat d'une panne du réseau ou d'une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.
- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, y compris la défaillance du site ou du système de stockage non privilégié, le site préféré libère l'état RPO=0 et reprend le traitement des E/S de lecture et d'écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration des services

Après résolution d'une panne, par exemple lors de la restauration de la connectivité site à site ou de la mise sous tension d'un système défaillant, les terminaux de synchronisation active SnapMirror détectent automatiquement la présence d'une relation de réplication défectueuse et la raverront à l'état RPO=0. Une fois la réplication synchrone rétablie, les chemins défaillants se reconnectent.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement. Cela dépend de l'application et de la façon dont elle est configurée et, en général, de telles tâches peuvent être facilement

automatisées. La fonctionnalité ONTAP elle-même est dotée d'une fonctionnalité d'autorétablissement et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0.

Basculement manuel

La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

Configuration du stockage sur les systèmes ASA r2

Présentation

NetApp ASA r2 est une solution puissante et simplifiée destinée aux clients SAN exécutant des workloads stratégiques. L'association de la plateforme ASA r2 qui exécute les solutions de stockage ONTAP et de Microsoft SQL Server permet de concevoir un stockage de base de données d'entreprise capable de répondre aux exigences des applications les plus exigeantes.

Les plateformes ASA suivantes sont classifiées comme des systèmes ASA r2 prenant en charge tous les protocoles SAN (iSCSI, FC, NVMe/FC, NVMe/TCP). Les protocoles iSCSI, FC, NVMe/FC et NVMe/TCP prennent en charge une architecture actif-actif symétrique pour les chemins d'accès multiples, de sorte que tous les chemins entre les hôtes et le stockage soient actifs/optimisés :

- ASA A1K
- ASA A90
- ASA A70
- ASA A50
- ASA A30
- ASA A20

Pour plus d'informations, voir "NetApp ASA"

Pour optimiser une solution SQL Server sur ONTAP, il est nécessaire de comprendre le modèle et les caractéristiques d'E/S SQL Server. Une infrastructure de stockage bien conçue pour une base de données SQL Server doit répondre aux besoins de performances de SQL Server, tout en assurant une gestion maximale de l'infrastructure dans son ensemble. Une bonne disposition du stockage permet également de réussir le déploiement initial et d'assurer une croissance progressive de l'environnement à mesure que l'entreprise se développe.

Conception du stockage des données

Microsoft recommande de placer les fichiers de données et les fichiers journaux sur des lecteurs distincts. Pour les applications qui mettent à jour et demandent simultanément des données, le fichier journal est très gourmand en écriture et le fichier de données (selon votre application) consomme beaucoup de ressources en lecture/écriture. Pour la récupération des données, le fichier journal n'est pas nécessaire. Par conséquent, les demandes de données peuvent être satisfaites à partir du fichier de données placé sur son propre disque.

Lorsque vous créez une nouvelle base de données, Microsoft recommande de spécifier des disques distincts

pour les données et les journaux. Pour déplacer des fichiers après la création de la base de données, la base de données doit être mise hors ligne. Pour plus d'informations sur les recommandations de Microsoft, consultez la section "Placez les fichiers de données et les fichiers journaux sur des lecteurs distincts".

Considérations relatives à l'unité de stockage

Unité de stockage dans ASA désigne un LUN pour les hôtes SCSI/FC ou un namespace NVMe pour les hôtes NVMe. En fonction du protocole pris en charge, vous serez invité à créer une LUN, un espace de noms NVMe ou les deux. Avant de créer une unité de stockage pour le déploiement de la base de données, il est important de comprendre comment le modèle et les caractéristiques d'E/S SQL Server varient en fonction de la charge de travail et des exigences de sauvegarde et de restauration. Consultez les recommandations NetApp suivantes pour l'unité de stockage :

- Évitez de partager la même unité de stockage entre plusieurs serveurs SQL Server s'exécutant sur le même hôte afin d'éviter une gestion complexe. Dans le cas d'une exécution de plusieurs instances SQL Server sur le même hôte, sauf si vous êtes proche de la limite de l'unité de stockage sur un nœud, évitez le partage et disposez plutôt d'une unité de stockage distincte par instance et par hôte pour faciliter la gestion des données.
- Utilisez des points de montage NTFS au lieu de lettres de lecteur pour dépasser la limite de 26 lettres de lecteur dans Windows.
- Désactivez les planifications d'instantanés et les stratégies de conservation. Utilisez plutôt SnapCenter pour coordonner les copies Snapshot de l'unité de stockage des données SQL Server.
- Placez les bases de données système SQL Server sur une unité de stockage dédiée.
- Tempdb est une base de données système utilisée par SQL Server comme espace de travail temporaire, en particulier pour les opérations DBCC CHECKDB exigeantes en E/S. Par conséquent, placez cette base de données sur une unité de stockage dédiée. Dans les grands environnements dans lesquels le nombre d'unités de stockage est un défi, vous pouvez consolider tempdb avec des bases de données système dans la même unité de stockage après une planification minutieuse. La protection des données pour tempdb n'est pas une priorité élevée car cette base de données est recréée à chaque redémarrage de SQL Server.
- Placez les fichiers de données utilisateur (`.mdf`) sur une unité de stockage distincte car il s'agit de charges de travail en lecture/écriture aléatoires. Il est courant de créer des sauvegardes du journal de transactions plus fréquemment que les sauvegardes de bases de données. Pour cette raison, placez les fichiers journaux (`.ldf`de transactions) sur une unité de stockage distincte ou un fichier VMDK à partir des fichiers de données afin que des planifications de sauvegarde indépendantes puissent être créées pour chaque. Cette séparation isole également les E/S d'écriture séquentielle des fichiers journaux des E/S de lecture/écriture aléatoires des fichiers de données et améliore considérablement les performances de SQL Server.
- Assurez-vous que les fichiers de base de données utilisateur et le répertoire des journaux pour stocker la sauvegarde des journaux se trouvent sur une unité de stockage distincte afin d'empêcher la règle de rétention d'écraser les snapshots lorsqu'ils sont utilisés avec la fonction SnapMirror avec la règle de coffre-fort.
- Ne mélangez pas les fichiers de base de données et les fichiers non liés à la base de données, tels que les fichiers de recherche en texte intégral, sur la même unité de stockage.
- Le placement de fichiers secondaires de base de données (dans le cadre d'un groupe de fichiers) sur une unité de stockage distincte améliore les performances de la base de données SQL Server. Cette séparation est valide uniquement si le fichier de la base de données `.mdf` ne partage pas son unité de stockage avec d'autres `.mdf` fichiers.
- Lors du formatage du disque à l'aide du gestionnaire de disques dans le serveur Windows, assurez-vous que la taille de l'unité d'allocation est définie sur 64 Ko pour la partition.

- Ne placez pas de bases de données utilisateur ou de bases de données système sur une unité de stockage hébergeant des points de montage.
- Voir la "Microsoft Windows et MPIO natif conformément aux meilleures pratiques ONTAP pour les SAN modernes" Pour appliquer la prise en charge des chemins d'accès multiples sur Windows aux périphériques iSCSI dans les propriétés MPIO.
- Si vous utilisez une instance de cluster à basculement permanent, les bases de données utilisateur doivent être placées sur une unité de stockage partagée entre les nœuds de cluster de basculement du serveur Windows et les ressources de cluster de disques physiques sont affectées au groupe de clusters associé à l'instance SQL Server.

Fichiers de base de données et groupes de fichiers

Il est essentiel de placer correctement les fichiers de base de données SQL Server sur ONTAP lors de la phase de déploiement initiale. Vous bénéficiez ainsi de performances optimales, d'un temps de gestion de l'espace, de sauvegarde et de restauration qui peuvent être configurés pour répondre aux besoins de votre entreprise.

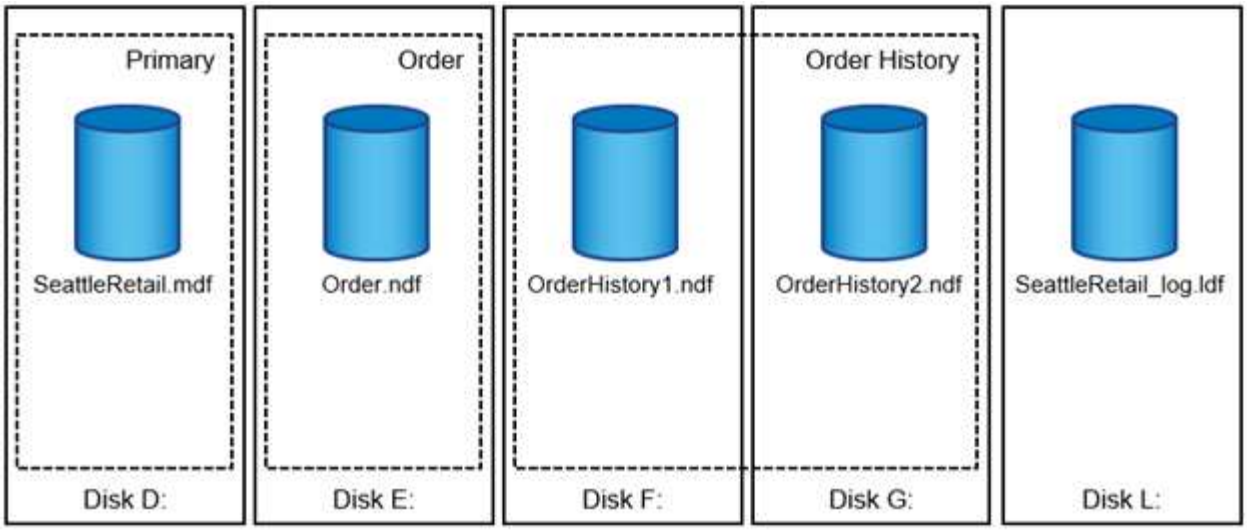
En théorie, SQL Server (64 bits) prend en charge 32,767 bases de données par instance et 524 272 To de taille de base de données, bien que l'installation standard comporte généralement plusieurs bases de données. Cependant, le nombre de bases de données que SQL Server peut gérer dépend de la charge et du matériel. Il n'est pas rare que des instances SQL Server hébergent des dizaines, des centaines, voire des milliers de petites bases de données.

Fichiers de base de données et groupe de fichiers

Chaque base de données se compose d'un ou plusieurs fichiers de données et d'un ou plusieurs fichiers journaux de transactions. Le journal de transactions stocke les informations sur les transactions de base de données et toutes les modifications de données effectuées par chaque session. Chaque fois que les données sont modifiées, SQL Server stocke suffisamment d'informations dans le journal de transactions pour annuler (revenir en arrière) ou rétablir (relire) l'action. Un journal de transactions SQL Server fait partie intégrante de la réputation de SQL Server en matière d'intégrité et de robustesse des données. Le journal de transactions est essentiel aux capacités d'atomicité, de cohérence, d'isolation et de durabilité (ACIDE) de SQL Server. SQL Server écrit dans le journal de transactions dès qu'une modification de la page de données se produit. Chaque instruction Data manipulation Language (DML) (par exemple, Select, INSERT, Update ou DELETE) est une transaction complète, et le journal de transactions s'assure que l'opération basée sur l'ensemble a lieu, en s'assurant de l'atomicité de la transaction.

Chaque base de données possède un fichier de données primaire, qui, par défaut, possède l'extension .mdf. En outre, chaque base de données peut avoir des fichiers de base de données secondaires. Ces fichiers, par défaut, ont des extensions .ndf.

Tous les fichiers de base de données sont regroupés en groupes de fichiers. Un groupe de fichiers est l'unité logique, qui simplifie l'administration de la base de données. Ils permettent de séparer le placement d'objets logiques des fichiers de base de données physiques. Lorsque vous créez les tables d'objets de base de données, vous spécifiez dans quel groupe de fichiers elles doivent être placées sans vous soucier de la configuration du fichier de données sous-jacent.



La possibilité de placer plusieurs fichiers de données dans le groupe de fichiers vous permet de répartir la charge entre les différents périphériques de stockage, ce qui contribue à améliorer les performances d'E/S du système. En revanche, le journal de transactions ne bénéficie pas des multiples fichiers car SQL Server écrit dans le journal de transactions de manière séquentielle.

La séparation entre le placement d'objets logiques dans les groupes de fichiers et les fichiers de base de données physiques vous permet d'affiner la disposition des fichiers de base de données, en tirant le meilleur parti du sous-système de stockage. Le nombre de fichiers de données prenant en charge une charge de travail donnée peut varier en fonction des besoins pour prendre en charge les exigences d'E/S et la capacité prévue, sans affecter l'application. Ces variations dans la disposition de la base de données sont transparentes pour les développeurs d'applications, qui placent les objets de base de données dans les groupes de fichiers plutôt que dans les fichiers de base de données.

NetApp recommande d'éviter l'utilisation du groupe de fichiers principal pour tout autre objet que les objets système. La création d'un groupe de fichiers distinct ou d'un ensemble de groupes de fichiers pour les objets utilisateur simplifie l'administration de la base de données et la reprise après incident, en particulier dans le cas de bases de données volumineuses.

Initialisation du fichier d'instance de base de données

Vous pouvez spécifier la taille initiale du fichier et les paramètres de croissance automatique au moment de la création de la base de données ou de l'ajout de nouveaux fichiers à une base de données existante. SQL Server utilise un algorithme de remplissage proportionnel lors du choix du fichier de données dans lequel il doit écrire des données. Elle écrit une quantité de données proportionnellement à l'espace libre disponible dans les fichiers. Plus l'espace libre dans le fichier est important, plus il traite d'écritures.

NetApp recommande que tous les fichiers d'un seul groupe de fichiers aient les mêmes paramètres de taille initiale et de croissance automatique, avec la taille de croissance définie en mégaoctets plutôt qu'en pourcentages. Cela permet à l'algorithme de remplissage proportionnel d'équilibrer uniformément les activités d'écriture entre les fichiers de données.

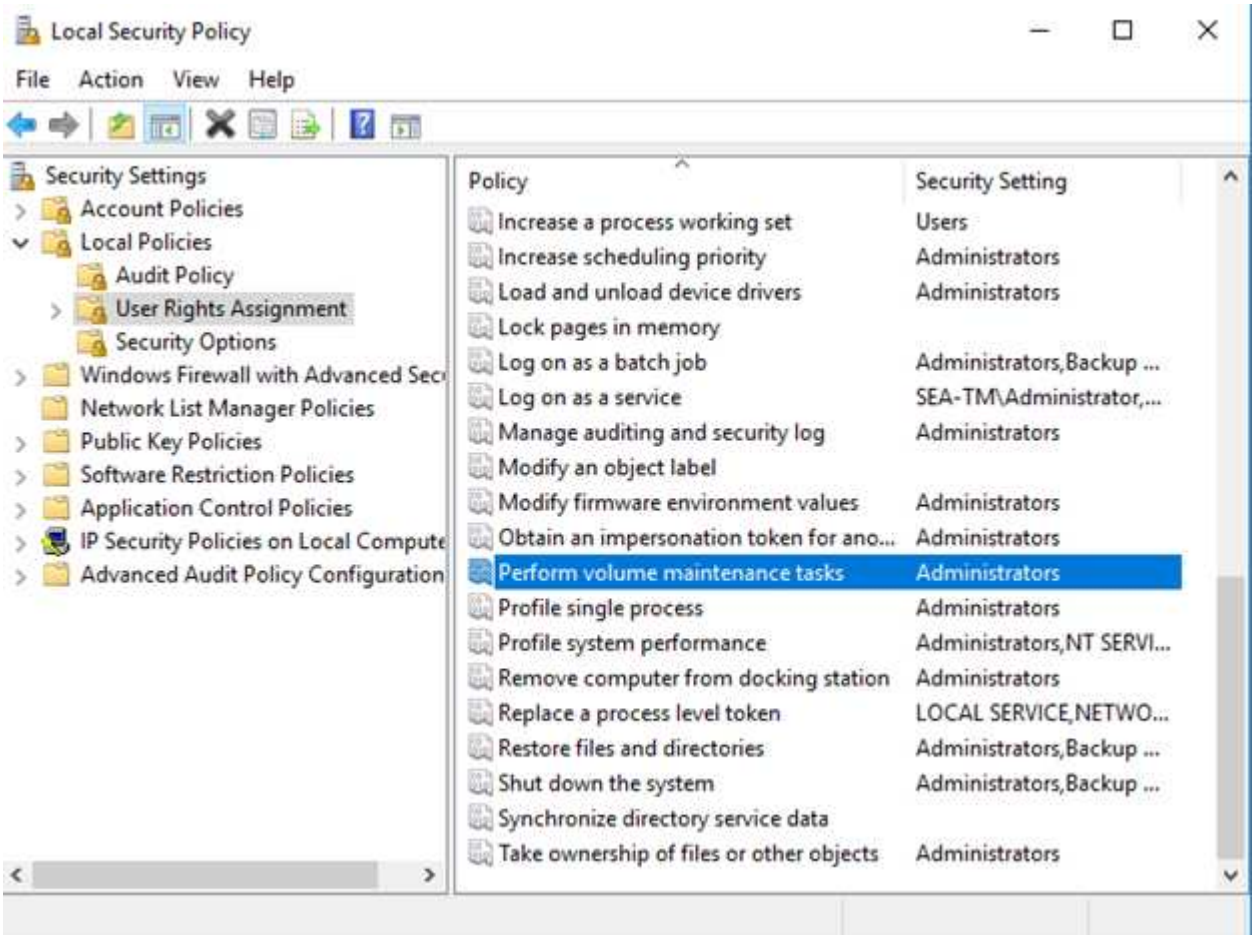
Chaque fois que SQL Server augmente la taille des fichiers, il remplit l'espace nouvellement alloué avec des zéros. Ce processus bloque toutes les sessions qui doivent écrire dans le fichier correspondant ou, en cas de croissance du journal de transactions, générer des enregistrements de journal de transactions.

SQL Server met toujours à zéro le journal de transactions et ce comportement ne peut pas être modifié. Toutefois, vous pouvez contrôler si les fichiers de données sont mis à zéro en activant ou en désactivant

l'initialisation instantanée des fichiers. L'activation de l'initialisation instantanée des fichiers permet d'accélérer la croissance des fichiers de données et de réduire le temps nécessaire à la création ou à la restauration de la base de données.

Un petit risque de sécurité est associé à l'initialisation instantanée des fichiers. Lorsque cette option est activée, les parties non allouées du fichier de données peuvent contenir des informations provenant de fichiers OS précédemment supprimés. Les administrateurs de base de données peuvent examiner ces données.

Vous pouvez activer l'initialisation instantanée des fichiers en ajoutant l'autorisation sa_MANAGE_VOLUME_NAME, également appelée « effectuer une tâche de maintenance de volume » au compte de démarrage SQL Server. Vous pouvez le faire sous l'application de gestion des stratégies de sécurité locales (secpol.msc), comme indiqué dans la figure suivante. Ouvrez les propriétés de l'autorisation "effectuer une tâche de maintenance de volume" et ajoutez le compte de démarrage SQL Server à la liste des utilisateurs.



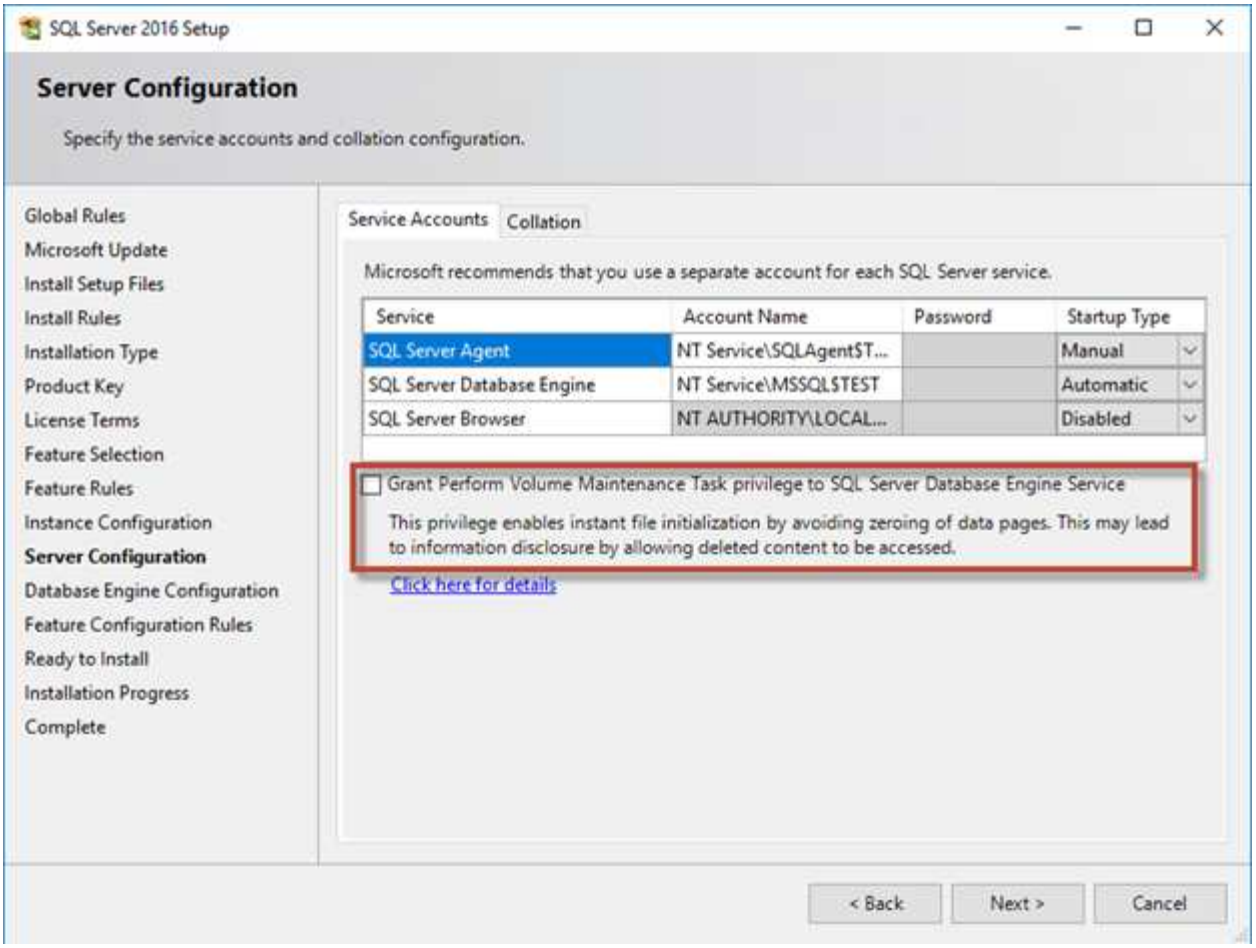
Pour vérifier si l'autorisation est activée, vous pouvez utiliser le code de l'exemple suivant. Ce code définit deux indicateurs de suivi qui forcent SQL Server à écrire des informations supplémentaires dans le journal d'erreurs, à créer une petite base de données et à lire le contenu du journal.

```
DBCC TRACEON(3004,3605,-1)
GO
CREATE DATABASE DelMe
GO
EXECUTE sp_readerrorlog
GO
DROP DATABASE DelMe
GO
DBCC TRACEOFF(3004,3605,-1)
GO
```

Lorsque l'initialisation instantanée des fichiers n'est pas activée, le journal d'erreurs SQL Server indique que SQL Server met à zéro le fichier de données mdf en plus de mettre à zéro le fichier journal ldf, comme indiqué dans l'exemple suivant. Lorsque l'initialisation instantanée des fichiers est activée, elle affiche uniquement la remise à zéro du fichier journal.

	LogDate	ProcessInfo	Text
365	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 flush delta counts.
366	2017-02-09 08:10:07.660	spid53	Ckpt dbid 3 logging active xact info.
367	2017-02-09 08:10:07.750	spid53	Ckpt dbid 3 phase 1 ended (8)
368	2017-02-09 08:10:07.750	spid53	About to log Checkpoint end.
369	2017-02-09 08:10:07.880	spid53	Ckpt dbid 3 complete
370	2017-02-09 08:10:08.130	spid53	Starting up database 'DelMe'.
371	2017-02-09 08:10:08.150	spid53	FixupLog Tail(progress) zeroing C:\Program Files\Micros
372	2017-02-09 08:10:08.160	spid53	Zeroing C:\Program Files\Microsoft SQL Server\MSSQ
373	2017-02-09 08:10:08.170	spid53	Zeroing completed on C:\Program Files\Microsoft SQL
374	2017-02-09 08:10:08.710	spid53	Ckpt dbid 6 started
375	2017-02-09 08:10:08.710	spid53	About to log Checkpoint begin.

La tâche effectuer une maintenance de volume est simplifiée dans SQL Server 2016 et est fournie ultérieurement en option pendant le processus d'installation. Cette figure affiche l'option permettant d'accorder au service du moteur de base de données SQL Server le privilège d'effectuer la tâche de maintenance du volume.



Une autre option de base de données importante qui contrôle la taille des fichiers de base de données est la fonction de transmission automatique. Lorsque cette option est activée, SQL Server réduit régulièrement les fichiers de base de données, réduit leur taille et libère de l'espace dans le système d'exploitation. Cette opération consomme beaucoup de ressources et est rarement utile car les fichiers de base de données augmentent à nouveau après l'arrivée de nouvelles données dans le système. La fonction Autohrink ne doit pas être activée sur la base de données.

Répertoire du journal

Le répertoire du journal est spécifié dans SQL Server pour stocker les données de sauvegarde du journal de transactions au niveau de l'hôte. Si vous utilisez SnapCenter pour sauvegarder les fichiers journaux, chaque hôte SQL Server utilisé par SnapCenter doit disposer d'un répertoire de journaux hôte configuré pour effectuer des sauvegardes de journaux.

Placez le répertoire des journaux sur une unité de stockage dédiée. La quantité de données dans le répertoire du journal hôte dépend de la taille des sauvegardes et du nombre de jours pendant lesquels les sauvegardes sont conservées. SnapCenter n'autorise qu'un seul répertoire de journaux hôte par hôte SQL Server. Vous pouvez configurer les répertoires de journaux hôtes dans SnapCenter -→ hôte -→ configurer le plug-in.

NetApp recommande ce qui suit pour un répertoire de journaux hôte :

- Assurez-vous que le répertoire du journal de l'hôte n'est partagé par aucun autre type de données pouvant potentiellement corrompre les données du snapshot de sauvegarde.
- Créez le répertoire des journaux hôtes sur une unité de stockage dédiée dans laquelle SnapCenter copie les journaux de transactions.
- Si vous utilisez une instance de cluster de basculement toujours en service, l'unité de stockage utilisée pour le répertoire des journaux de l'hôte doit être une ressource de disque de cluster dans le même groupe que l'instance SQL Server sauvegardée dans SnapCenter.

Protection des données

Les stratégies de sauvegarde des bases de données doivent être basées sur des exigences métier identifiées, et non sur des capacités théoriques. En combinant la technologie Snapshot de ONTAP et en exploitant les API de Microsoft SQL Server, vous pouvez effectuer rapidement des sauvegardes cohérentes au niveau des applications, quelle que soit la taille des bases de données utilisateur. Pour une gestion des données plus avancée ou scale-out, NetApp propose SnapCenter.

SnapCenter

SnapCenter est le logiciel NetApp de protection des données pour les applications d'entreprise. Les bases de données SQL Server peuvent être protégées rapidement et facilement grâce au plug-in SnapCenter pour SQL Server et aux opérations du système d'exploitation gérées par le plug-in SnapCenter pour Microsoft Windows.

L'instance SQL Server peut être une instance de cluster d'installation autonome ou de basculement, ou elle peut être toujours sur le groupe de disponibilité. Le résultat est que depuis une fenêtre unique, les bases de données peuvent être protégées, clonées et restaurées à partir d'une copie principale ou secondaire. SnapCenter peut gérer les bases de données SQL Server à la fois sur site, dans le cloud et dans des configurations hybrides.des copies de bases de données peuvent également être créées en quelques minutes sur l'hôte original ou alternatif à des fins de développement ou de reporting.

SQL Server nécessite également une coordination entre le système d'exploitation et le stockage pour s'assurer que les données correctes sont présentes dans les snapshots au moment de la création. Dans la plupart des cas, la seule méthode sûre pour ce faire est SnapCenter ou T-SQL. Les snapshots créés sans cette coordination supplémentaire peuvent ne pas être récupérables de manière fiable.

Pour plus d'informations sur le plug-in SQL Server pour SnapCenter, reportez-vous à la section "Tr-4714 : guide des meilleures pratiques pour SQL Server avec NetApp SnapCenter".

Protection de la base de données à l'aide de snapshots T-SQL

Dans SQL Server 2022, Microsoft a introduit des snapshots T-SQL qui permettent de réaliser des scripts et d'automatiser les opérations de sauvegarde. Au lieu d'effectuer des copies complètes, vous pouvez préparer la base de données pour les snapshots. Une fois la base de données prête pour la sauvegarde, vous pouvez utiliser les API REST de ONTAP pour créer des snapshots.

Voici un exemple de flux de travail de sauvegarde :

1. Figez une base de données à l'aide de la commande ALTER. La base de données est ainsi préparée pour un snapshot cohérent sur le stockage sous-jacent. Après le gel, vous pouvez dégeler la base de données et enregistrer le snapshot avec la commande BACKUP.

2. Réalisez des instantanés de plusieurs bases de données sur les unités de stockage simultanément avec les nouvelles commandes de GROUPE DE SAUVEGARDE et de SERVEUR DE SAUVEGARDE.
3. Si la charge de travail de la base de données est étendue à plusieurs unités de stockage, créez des groupes de cohérence pour simplifier la tâche de gestion. Le groupe de cohérence est un ensemble d'unités de stockage gérées comme une seule unité.
4. Effectuer des sauvegardes COMPLÈTES ou des sauvegardes COMPLÈTES COPY_ONLY. Ces sauvegardes sont également enregistrées dans msdb.
5. Effectuez une restauration instantanée à l'aide de sauvegardes de journaux effectuées avec l'approche de streaming standard après la sauvegarde COMPLÈTE des snapshots. Les sauvegardes différentielles en continu sont également prises en charge si nécessaire.

Pour en savoir plus, voir "Documentation Microsoft à connaître sur les snapshots T-SQL".

NetApp recommande d'utiliser SnapCenter pour créer des copies Snapshot. La méthode T-SQL décrite ci-dessus fonctionne également, mais SnapCenter offre une automatisation complète du processus de sauvegarde, de restauration et de clonage. Il effectue également une découverte pour s'assurer que les snapshots corrects sont créés.

Utilisez SnapCenter si les bases de données SQL Server sont hébergées sur un disque virtuel exécuté dans un environnement virtualisé VMware. Si les bases de données sont hébergées sur un serveur physique bare Metal ou sur un SCSI Guest-Connect, sauvegardez la base de données à l'aide de la commande T-SQL snapshot.

Reprise après incident

Reprise après incident

Les bases de données d'entreprise et les infrastructures applicatives ont souvent besoin d'une réplication pour se protéger contre les catastrophes naturelles ou les perturbations imprévues, avec un temps d'interruption minimal.

La fonction de réplication de groupe de disponibilité en continu de SQL Server peut constituer une excellente option, et NetApp offre des options pour intégrer la protection des données à la disponibilité continue. Toutefois, dans certains cas, il peut être intéressant d'opter pour la technologie de réplication ONTAP en utilisant les options suivantes.

SnapMirror

La technologie SnapMirror offre une solution d'entreprise rapide et flexible pour la réplication de données sur des réseaux LAN et WAN. La technologie SnapMirror transfère uniquement les blocs de données modifiés vers la destination après la création du miroir initial, ce qui réduit considérablement les besoins en bande passante réseau. Il peut être configuré en mode synchrone ou asynchrone. La réplication synchrone SnapMirror dans NetApp ASA est configurée à l'aide de la synchronisation active SnapMirror.

Synchronisation active SnapMirror

Pour de nombreux clients, la continuité de l'activité ne se limite pas à posséder une copie distante des données. Il est donc impératif de pouvoir utiliser rapidement ces données dans NetApp ONTAP à l'aide de la synchronisation active SnapMirror

Avec la synchronisation active SnapMirror, deux systèmes ONTAP différents conservent des copies

indépendantes de vos données LUN, mais fonctionnent ensemble pour présenter une seule instance de ce LUN. Du point de vue de l'hôte, il s'agit d'une entité LUN unique. La synchronisation active SnapMirror est prise en charge pour les LUN iSCSI/FC.

La synchronisation active SnapMirror peut assurer une réplication avec un objectif de point de récupération de 0 et elle est facile à implémenter entre deux clusters indépendants.une fois les deux copies de données synchronisées, les deux clusters n'ont besoin que de mettre en miroir les écritures. Lorsqu'une écriture a lieu sur un cluster, elle est répliquée sur l'autre. L'écriture est uniquement validée par l'hôte lorsque l'écriture est terminée sur les deux sites. En dehors de ce comportement de fractionnement de protocole, les deux clusters sont des clusters ONTAP normaux.

L'une des principales utilisations de SM-AS est la réplication granulaire. Parfois, vous ne souhaitez pas répliquer toutes les données en tant qu'unité unique ou vous devez pouvoir basculer sélectivement sur certains workloads.

Autre cas d'utilisation clé de la solution SM-as pour les opérations actives/actives : vous souhaitez que des copies de données entièrement exploitables soient disponibles sur deux clusters différents situés à deux emplacements différents avec des performances identiques et, si vous le souhaitez, vous n'avez pas besoin d'étendre le SAN sur plusieurs sites. Vos applications peuvent déjà s'exécuter sur les deux sites, à condition que l'application soit prise en charge, ce qui réduit l'objectif de délai de restauration global pendant les opérations de basculement.

SnapMirror

Voici les recommandations pour SnapMirror pour SQL Server :

- Utilisez la réplication synchrone avec SnapMirror Active Sync où la demande de restauration rapide des données est plus importante et des solutions asynchrones pour plus de flexibilité dans le RPO.
- Si vous utilisez SnapCenter pour sauvegarder des bases de données et répliquer des snapshots sur un cluster distant, ne planifiez pas les mises à jour SnapMirror à partir des contrôleurs à des fins de cohérence. Activez plutôt les mises à jour SnapMirror depuis SnapCenter pour mettre à jour SnapMirror une fois la sauvegarde complète ou la sauvegarde du journal terminée.
- Équilibrez les unités de stockage contenant des données SQL Server sur différents nœuds du cluster afin de permettre à tous les nœuds du cluster de partager l'activité de réplication SnapMirror. Cette distribution optimise l'utilisation des ressources du nœud.

Pour plus d'informations sur SnapMirror, reportez-vous à la section "Tr-4015 : Guide de configuration et des meilleures pratiques de SnapMirror pour ONTAP 9".

Synchronisation active SnapMirror

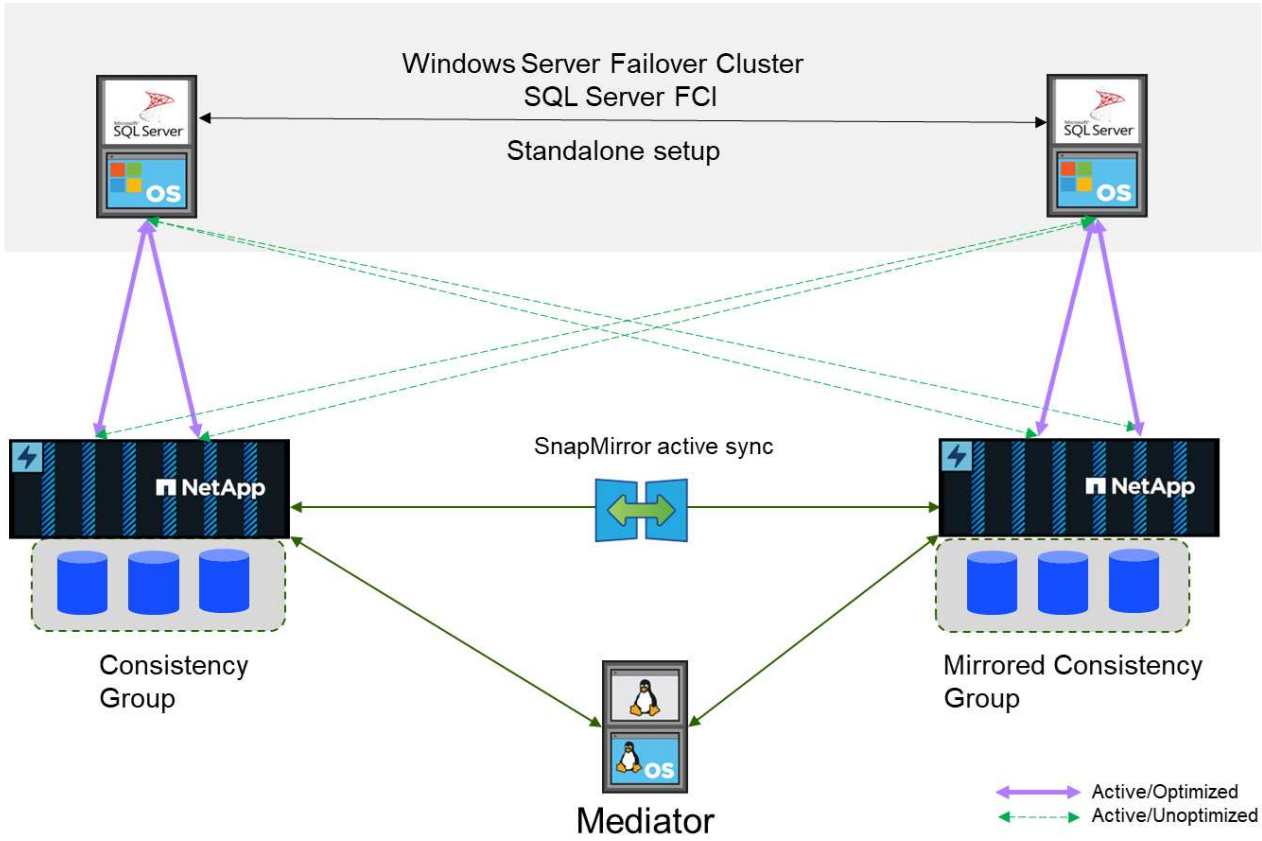
Présentation

La synchronisation active SnapMirror permet à chaque base de données et application SQL Server de continuer les opérations pendant les interruptions du stockage et du réseau, grâce à un basculement transparent du stockage, sans intervention manuelle.

La synchronisation active SnapMirror prend en charge une architecture actif-actif symétrique qui offre une réplication bidirectionnelle synchrone pour la continuité de l'activité et la reprise après incident. Il vous aide à protéger l'accès aux données pour les workloads SAN stratégiques avec un accès simultané en lecture et en écriture aux données dans plusieurs domaines à défaillance. Vous bénéficiez ainsi d'une continuité de l'activité et d'une réduction des temps d'indisponibilité en cas d'incident ou de panne système.

Les hôtes SQL Server accèdent au stockage via des LUN Fibre Channel (FC) ou iSCSI. Réplication entre chaque cluster hébergeant une copie des données répliquées. Étant donné que cette fonctionnalité est une réplication au niveau du stockage, les instances SQL Server exécutées sur des instances d'hôte autonome ou de cluster de basculement peuvent effectuer des opérations de lecture/écriture sur l'un ou l'autre des clusters. Pour les étapes de planification et de configuration, reportez-vous à la section "Documentation ONTAP sur la synchronisation active SnapMirror" .

Architecture SnapMirror active avec symétrie actif-actif



Réplication synchrone

En fonctionnement normal, chaque copie correspond à une réplique synchrone RPO=0 à tout moment, à une exception près. Si les données ne peuvent pas être répliquées, ONTAP exige de répliquer les données et de reprendre le traitement des E/S sur un site pendant que les LUN de l'autre site sont mises hors ligne.

Matériel de stockage

Contrairement à d'autres solutions de reprise après incident du stockage, la synchronisation active SnapMirror offre une flexibilité asymétrique de la plateforme. Le matériel de chaque site n'a pas besoin d'être identique. Cette fonctionnalité vous permet d'ajuster la taille du matériel utilisé pour prendre en charge la synchronisation active SnapMirror. Le système de stockage distant peut être identique au site principal s'il doit prendre en charge une charge de travail de production complète, mais si un incident entraîne une réduction des E/S, un système plus petit sur le site distant peut être plus économique.

Médiateur ONTAP

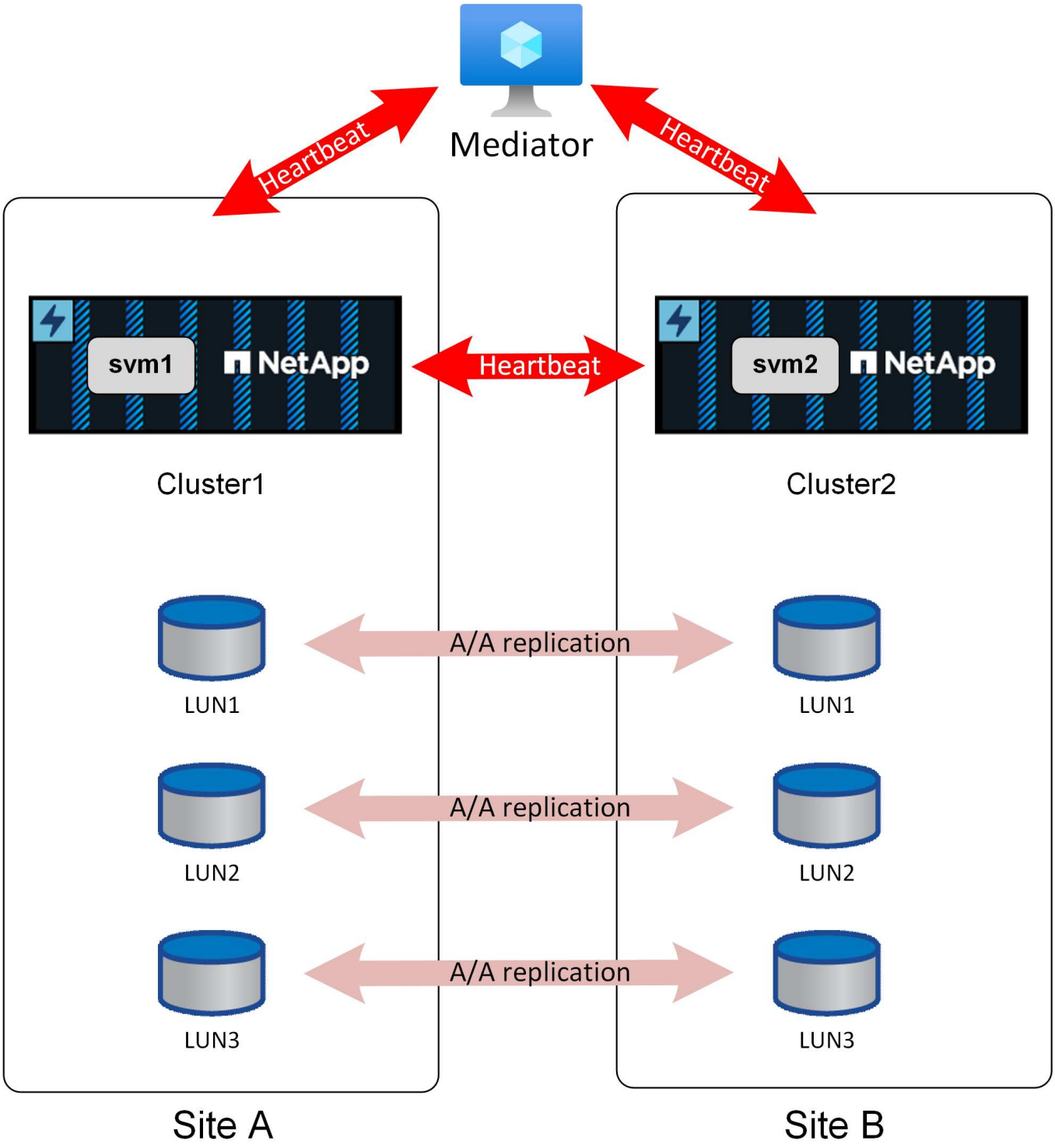
Le médiateur ONTAP est une application logicielle téléchargée depuis la prise en charge de NetApp et généralement déployée sur une petite machine virtuelle. Le médiateur ONTAP n'est pas un tiebreaker. Il s'agit

d’un canal de communication alternatif pour les deux clusters qui participent à la réplication SnapMirror active Sync. Les opérations automatisées sont dirigées par ONTAP sur la base des réponses reçues du partenaire via des relations directes et via le médiateur.

ONTAP Médiateur

Le médiateur est requis pour automatiser le basculement en toute sécurité. Dans l’idéal, elle serait placée sur un site tiers indépendant, mais elle peut toujours fonctionner pour la plupart des besoins si elle est en colocation avec l’un des clusters participant à la réplication.

Le médiateur n’est pas vraiment un casse-barre, bien que c’est effectivement la fonction qu’il fournit. Il ne prend aucune action ; il fournit plutôt un canal de communication alternatif pour la communication cluster à cluster.



Le principal défi lié au basculement automatisé est le problème des réseaux partagés, qui se pose en cas de perte de connectivité entre les deux sites. Que doit-on faire ? Vous ne voulez pas que deux sites différents se désignent comme les copies restantes des données, mais comment un seul site peut-il faire la différence entre la perte réelle du site opposé et l'incapacité à communiquer avec le site opposé ?

C'est là que le médiateur entre dans la photo. S'il est placé sur un troisième site, et chaque site a une connexion réseau distincte à ce site, alors vous avez un chemin supplémentaire pour chaque site pour valider l'état de santé de l'autre. Examinez à nouveau l'image ci-dessus et examinez les scénarios suivants.

- Que se passe-t-il si le médiateur échoue ou est inaccessible à partir d'un ou des deux sites ?
  - Les deux clusters peuvent toujours communiquer entre eux sur le même lien que celui utilisé pour les services de réplication.
  - Les données restent protégées avec un objectif de point de récupération de 0
- Que se passe-t-il si le site A tombe en panne ?
  - Le site B verra les deux canaux de communication tomber en panne.
  - Le site B prendra le contrôle des services de données, mais sans mise en miroir RPO=0
- Que se passe-t-il si le site B tombe en panne ?
  - Le site A verra les deux canaux de communication tomber en panne.
  - Le site A prend le relais des services de données, mais sans mise en miroir avec un objectif de point de récupération de 0

Il y a un autre scénario à prendre en compte : la perte du lien de réplication des données. En cas de perte de la liaison de réplication entre les sites, la mise en miroir avec un objectif de point de récupération de 0 sera évidemment impossible. Que devrait-on alors se passer ?

Ceci est contrôlé par le statut du site préféré. Dans une relation SM-AS, l'un des sites est secondaire à l'autre. Cela n'a aucun effet sur les opérations normales, et tout accès aux données est symétrique. Toutefois, si la réplication est interrompue, le nœud devra être rompu pour reprendre les opérations. Par conséquent, le site privilégié continuera les opérations sans mise en miroir et le site secondaire arrêtera le traitement des E/S jusqu'à ce que la communication de réplication soit restaurée.

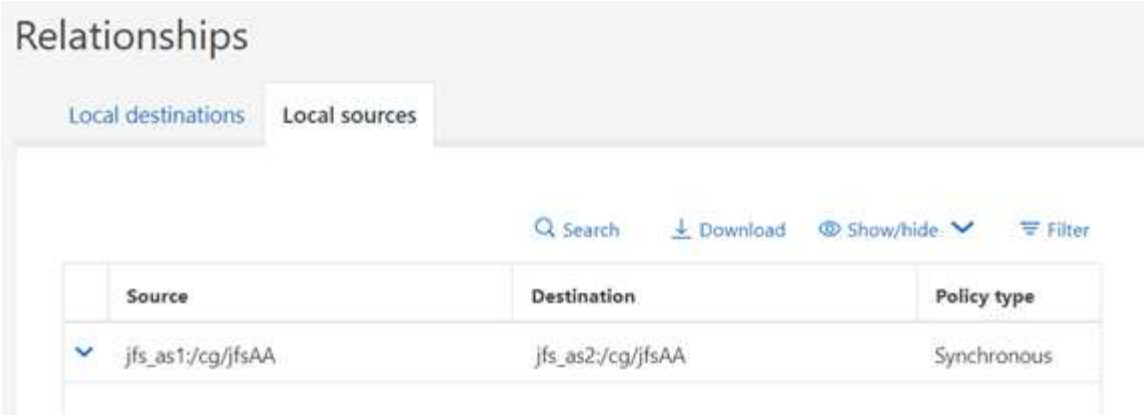
Site préféré

Le comportement de la synchronisation active SnapMirror est symétrique, avec une exception importante : la configuration du site préféré.

La synchronisation active SnapMirror considère un site comme la « source » et l'autre comme la « destination ». Cela implique une relation de réplication unidirectionnelle, mais cela ne s'applique pas au comportement d'E/S. La réplication est bidirectionnelle et symétrique. Les temps de réponse d'E/S sont identiques de part et d'autre du miroir.

La `source` désignation est le contrôle du site préféré. En cas de perte du lien de réplication, les chemins de LUN sur la copie source continueront à transmettre des données tandis que les chemins de LUN sur la copie de destination deviendront indisponibles jusqu'à ce que la réplication soit rétablie et que SnapMirror repasse à l'état synchrone. Les chemins reprennent alors le service des données.

La configuration source/destination peut être affichée via SystemManager :



Ou sur l'interface de ligne de commande :

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                      Destination Path: jfs_as2:/cg/jfsAA
                    Relationship Type: XDP
                Relationship Group Type: consistencygroup
                    SnapMirror Schedule: -
                 SnapMirror Policy Type: automated-failover-duplex
                      SnapMirror Policy: AutomatedFailOverDuplex
                            Tries Limit: -
                      Throttle (KB/sec): -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

La clé est que la source est le SVM sur le cluster1. Comme mentionné ci-dessus, les termes « source » et « destination » ne décrivent pas le flux des données répliquées. Les deux sites peuvent traiter une écriture et la répliquer sur le site opposé. En effet, les deux grappes sont des sources et des destinations. La désignation d'un cluster comme source contrôle simplement le cluster qui survit en tant que système de stockage en lecture/écriture en cas de perte du lien de réplication.

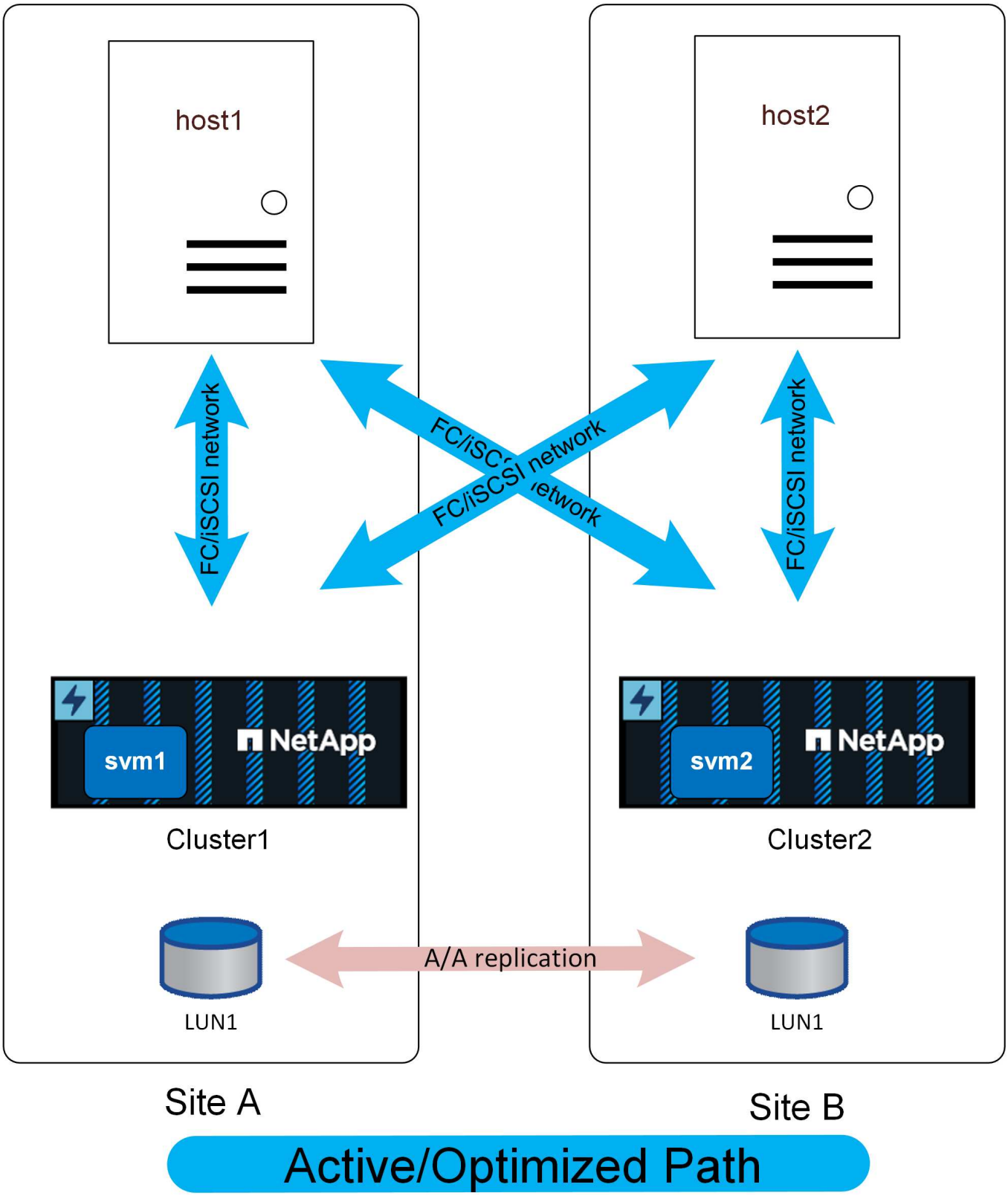
Topologie réseau

Accès uniforme

Un réseau d'accès uniforme signifie que les hôtes peuvent accéder aux chemins sur les deux sites (ou domaines de défaillance au sein du même site).

L'une des caractéristiques importantes de SM-AS est la capacité de configurer les systèmes de stockage pour savoir où se trouvent les hôtes. Lorsque vous mappez les LUN sur un hôte donné, vous pouvez indiquer si elles sont proximales ou non à un système de stockage donné.

Les systèmes NetApp ASA proposent des chemins d'accès multiples actif-actif sur tous les chemins d'accès à un cluster. Cela s'applique également aux configurations SM-AS.



Avec un accès uniforme, l'E/S traverserait le WAN. Il s'agit d'un cluster en réseau entièrement maillé, ce qui peut être souhaitable ou non pour tous les cas d'utilisation.

Si les deux sites étaient distants de 100 mètres avec une connectivité à fibre optique, il ne devrait pas y avoir de latence supplémentaire détectable traversant le WAN, mais si les sites étaient éloignés, les performances

de lecture seraient affectées sur les deux sites. ASA avec un réseau d'accès non uniforme serait une option pour bénéficier des avantages de ASA en termes de coûts et de fonctionnalités sans encourir de pénalités de latence intersites ou utiliser la fonction de proximité des hôtes pour autoriser l'accès en lecture/écriture locale des deux sites.

ASA avec SM-as dans une configuration à faible latence offre deux avantages intéressants. Tout d'abord, elle double *les performances de n'importe quel hôte, car les E/S peuvent être traitées par deux fois plus de contrôleurs en utilisant deux fois plus de chemins. Ensuite, dans un environnement à site unique, elle offre une disponibilité extrême, car l'intégralité du système de stockage peut être perdue sans interrompre l'accès aux hôtes.

Paramètres de proximité

La proximité fait référence à une configuration par cluster qui indique qu'un WWN d'hôte ou un ID d'initiateur iSCSI appartient à un hôte local. Il s'agit d'une deuxième étape facultative de configuration de l'accès aux LUN.

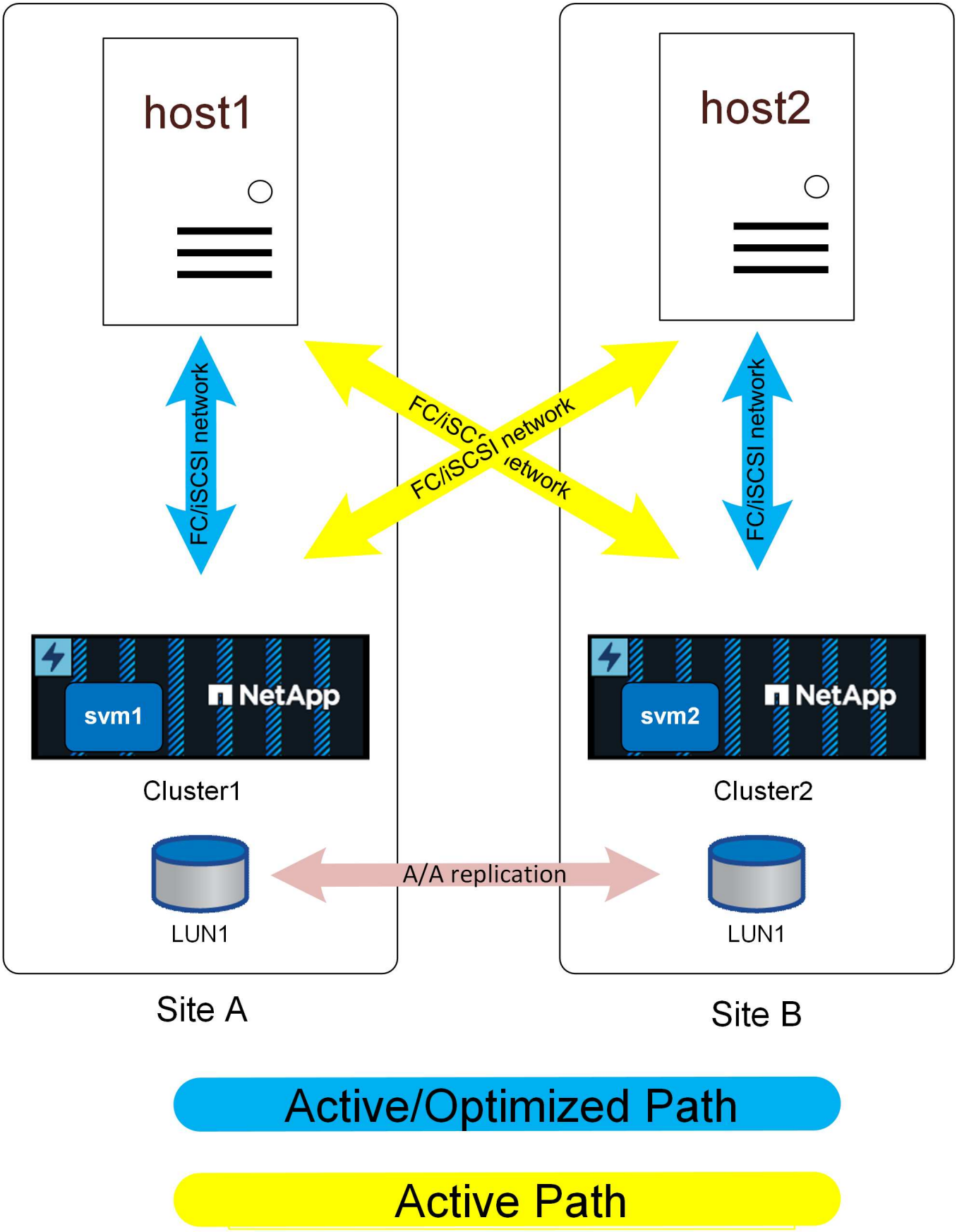
La première étape correspond à la configuration habituelle du groupe initiateur. Chaque LUN doit être mappée sur un groupe initiateur qui contient les ID WWN/iSCSI des hôtes devant accéder à cette LUN. Cela contrôle quel hôte a *accès* à un LUN.

La deuxième étape facultative consiste à configurer la proximité de l'hôte. Cela ne contrôle pas l'accès, il contrôle *Priority*.

Par exemple, un hôte du site A peut être configuré pour accéder à une LUN protégée par la synchronisation active SnapMirror. Le SAN étant étendu entre les sites, les chemins d'accès sont disponibles pour cette LUN via le stockage sur le site A ou le stockage sur le site B.

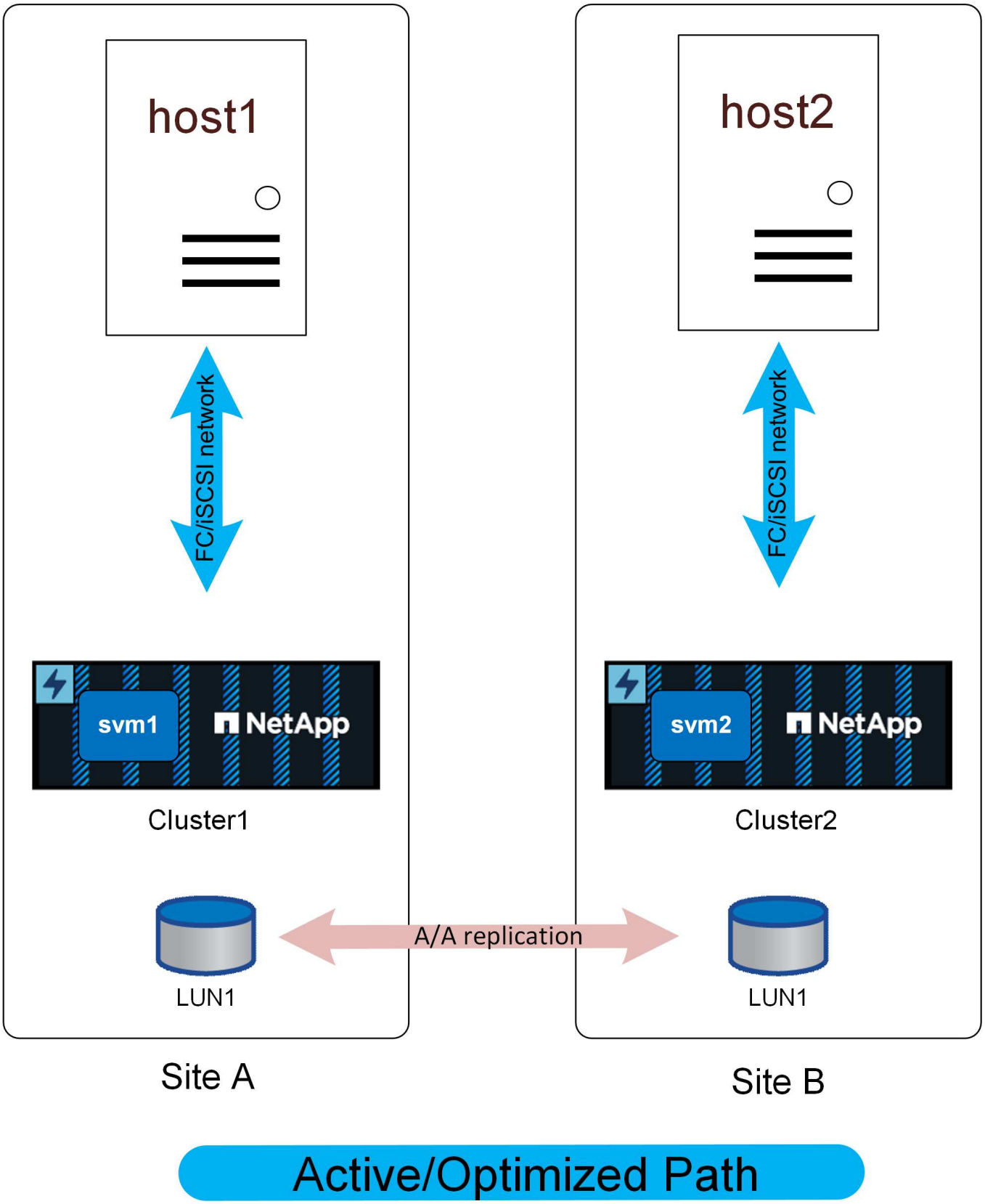
Sans paramètres de proximité, cet hôte utilisera les deux systèmes de stockage de la même manière, car les deux systèmes de stockage annonceront des chemins actifs/optimisés. Si la latence SAN et/ou la bande passante entre les sites est limitée, il se peut que cela ne soit pas désirable, et vous pouvez vous assurer que, pendant le fonctionnement normal, chaque hôte utilise de préférence des chemins vers le système de stockage local. Cette configuration s'effectue en ajoutant l'ID WWN/iSCSI de l'hôte au cluster local en tant qu'hôte proximal. Cette opération peut être effectuée à partir de l'interface de ligne de commande ou de SystemManager.

Les chemins s'affichent comme indiqué ci-dessous lorsque la proximité de l'hôte a été configurée.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
NetApp
svm2
NetApp
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

Accès non uniforme

La mise en réseau à accès non uniforme signifie que chaque hôte n’a accès qu’aux ports du système de stockage local. Le SAN n’est pas étendu sur les sites (ou les domaines de défaillance au sein du même site).



Le principal avantage de cette approche est la simplicité du SAN : vous n'avez plus besoin d'étendre un SAN sur le réseau. Certains clients ne disposent pas d'une connectivité à faible latence suffisante entre les sites, ou n'ont pas l'infrastructure nécessaire pour acheminer le trafic SAN FC sur un réseau intersite.

L'inconvénient de l'accès non uniforme est que certains scénarios de défaillance, notamment la perte du lien de réplication, entraînent la perte de l'accès au stockage par certains hôtes. En cas de perte de la connectivité du stockage local, les applications qui s'exécutent en tant qu'instances uniques, telles qu'une base de données non en cluster et qui ne s'exécute intrinsèquement que sur un hôte unique sur un montage donné, échouent. Les données seraient toujours protégées, mais le serveur de base de données n'aurait plus accès. Il doit être redémarré sur un site distant, de préférence par le biais d'un processus automatisé. Par exemple, VMware HA peut détecter une situation de tous les chemins d'accès sur un serveur et redémarrer une machine virtuelle sur un autre serveur sur lequel les chemins d'accès sont disponibles.

En revanche, une application en cluster telle qu'Oracle RAC peut fournir un service qui est disponible simultanément sur deux sites différents. La perte d'un site ne signifie pas la perte du service applicatif dans son ensemble. Les instances restent disponibles et s'exécutent sur le site survivant.

Dans de nombreux cas, la surcharge liée à la latence supplémentaire qu'une application accède au système de stockage via une liaison site à site ne serait pas acceptable. Cela signifie que l'amélioration de la disponibilité des réseaux uniformes est minime, car la perte de stockage sur un site entraînerait la nécessité de fermer les services sur ce site défaillant.

Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

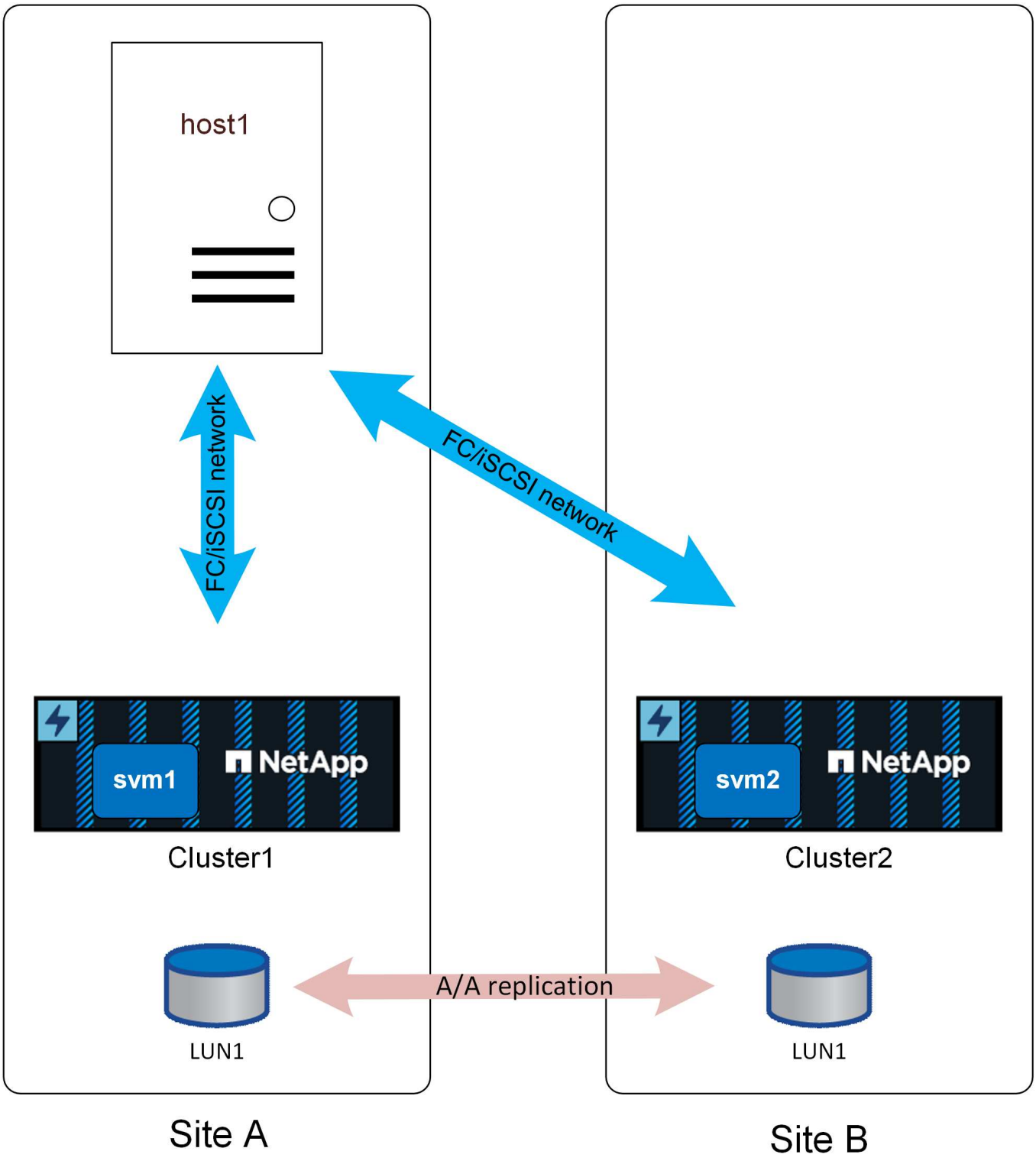
Présentation

SQL Server peut être configuré pour fonctionner avec la synchronisation active SnapMirror de plusieurs façons. La bonne réponse dépend de la connectivité réseau disponible, des exigences de RPO et de la disponibilité.

Instance autonome de SQL Server

Les meilleures pratiques en matière de mise en page des fichiers et de configuration des serveurs sont les mêmes que celles recommandées dans "SQL Server sur ONTAP" la documentation.

Avec une configuration autonome, SQL Server ne peut être exécuté que sur un site. L'accès serait probablement "uniforme"utilisé.



Avec un accès uniforme, une panne de stockage sur l'un ou l'autre site n'interromprait pas les opérations de la base de données. Une défaillance complète du site incluant le serveur de base de données entraînerait, bien sûr, une panne.

Certains clients peuvent configurer un système d'exploitation s'exécutant sur le site distant avec une configuration SQL Server préconfigurée, mise à jour avec une version de build équivalente à celle de l'instance de production. Le basculement nécessite l'activation de cette instance autonome de SQL Server sur le site secondaire, la découverte des LUN et le démarrage de la base de données. Le processus complet peut être

automatisé avec l'applet de commande Windows PowerShell, car aucune opération n'est requise côté stockage.

"Non uniforme" l'accès peut également être utilisé, mais il en résulte une panne de la base de données si le système de stockage sur lequel était situé le serveur de base de données avait échoué car la base de données ne disposait pas de chemins d'accès au stockage. Cela peut toujours être acceptable dans certains cas. La synchronisation active SnapMirror offre toujours une protection des données avec un objectif de point de récupération de 0. En cas de défaillance du site, la copie restante est active et prête à reprendre les opérations en suivant la même procédure utilisée avec un accès uniforme que celle décrite ci-dessus.

Un processus de basculement simple et automatisé peut être configuré plus facilement grâce à l'utilisation d'un hôte virtualisé. Par exemple, si les fichiers de données SQL Server sont répliqués de manière synchrone sur le stockage secondaire avec un VMDK de démarrage, l'environnement complet peut être activé sur l'autre site en cas d'incident. Un administrateur peut activer manuellement l'hôte sur le site survivant ou automatiser le processus via un service tel que VMware HA.

Instance de cluster de basculement SQL Server

Les instances de basculement SQL Server peuvent également être hébergées sur un cluster de basculement Windows s'exécutant sur un serveur physique ou virtuel en tant que système d'exploitation invité. Cette architecture multi-hôtes fournit l'instance SQL Server et la résilience du stockage. Ce déploiement est utile dans les environnements très exigeants qui recherchent des processus de basculement robustes tout en maintenant des performances améliorées. Dans une configuration de cluster de basculement, lorsqu'un hôte ou un stockage primaire est affecté, SQL Services effectue un basculement vers l'hôte secondaire et, dans le même temps, le stockage secondaire est disponible pour transmettre les E/S. Aucun script d'automatisation ni aucune intervention de l'administrateur n'est nécessaire.

Scénarios d'échec

La planification d'une architecture complète d'applications de synchronisation active SnapMirror nécessite de comprendre comment les SM-AS répondront dans divers scénarios de basculement planifiés et non planifiés.

Pour les exemples suivants, supposons que le site A est configuré comme le site préféré.

Perte de la connectivité de réplication

Si la réplication SM-AS est interrompue, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A (site préféré)

Le résultat de l'échec de la liaison de réplication sur le site préféré sera une pause d'environ 15 secondes dans le traitement des E/S d'écriture, car ONTAP relance les opérations d'écriture répliquées avant de déterminer que la liaison de réplication est véritablement inaccessible. Au bout de 15 secondes, le site A du système reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Panne du système de stockage

Le résultat d'une défaillance du système de stockage est presque identique au résultat de la perte du lien de réplication. Le site survivant devrait subir une pause d'E/S d'environ 15 seconde. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Perte du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ». En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

La raison pour laquelle un médiateur est nécessaire pour un basculement automatisé sécurisé est parce qu'il serait autrement impossible à un cluster de stockage de déterminer si la perte de la communication bidirectionnelle était le résultat d'une panne du réseau ou d'une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.
- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, y compris la défaillance du site ou du système de stockage non privilégié, le site préféré libère l'état RPO=0 et reprend le traitement des E/S de lecture et d'écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration des services

Après résolution d'une panne, par exemple lors de la restauration de la connectivité site à site ou de la mise sous tension d'un système défaillant, les terminaux de synchronisation active SnapMirror détectent automatiquement la présence d'une relation de réplication défectueuse et la raverront à l'état RPO=0. Une fois la réplication synchrone rétablie, les chemins défaillants se reconnectent.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement. Cela dépend de l'application et de la façon dont elle est configurée et, en général, de telles tâches peuvent être facilement

automatisées. La fonctionnalité ONTAP elle-même est dotée d'une fonctionnalité d'autorétablissement et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0.

Basculement manuel

La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

Sécurité des données

La sécurisation d'un environnement de base de données SQL Server est un effort multidimensionnel qui va au-delà de la gestion de la base de données elle-même. ONTAP propose plusieurs fonctionnalités uniques conçues pour sécuriser l'aspect stockage de votre infrastructure de base de données.

Copies Snapshot

Les snapshots de stockage sont des répliques instantanées des données cible. La mise en œuvre d'ONTAP permet de définir diverses règles et de stocker jusqu'à 1024 copies Snapshot par volume. Les copies Snapshot dans ONTAP sont compactes. L'espace est uniquement utilisé lorsque le dataset d'origine change. Ils sont également en lecture seule. Un snapshot peut être supprimé, mais il ne peut pas être modifié.

Dans certains cas, les snapshots peuvent être programmés directement sur ONTAP. Dans d'autres cas, des logiciels tels que SnapCenter peuvent être requis pour orchestrer les opérations d'application ou de système d'exploitation avant de créer des snapshots. Quelle que soit l'approche la plus adaptée à votre charge de travail, une stratégie Snapshot agressive peut assurer la sécurité des données grâce à un accès facile et fréquent aux sauvegardes de tous les éléments, des LUN de démarrage aux bases de données stratégiques.

> Un volume flexible ONTAP, ou plus simplement, un volume n'est pas synonyme de LUN. Les volumes sont des conteneurs de gestion pour des données telles que des fichiers ou des LUN. Par exemple, une base de données peut être placée sur un jeu de bandes de 8 LUN, toutes les LUN étant contenues dans un seul volume.

Pour plus d'informations sur les instantanés, reportez-vous au "Documentation ONTAP."

Des snapshots inviolables

Depuis ONTAP 9.12.1, les snapshots ne sont pas uniquement en lecture seule et peuvent également être protégés contre la suppression accidentelle ou intentionnelle. Cette fonctionnalité s'appelle des snapshots inviolables. Une période de conservation peut être définie et appliquée via une règle Snapshot. Les snapshots obtenus ne peuvent pas être supprimés tant qu'ils n'ont pas atteint leur date d'expiration. Il n'y a pas de substitution administrative ou de centre de support.

Cela permet de s'assurer qu'un intrus, un collaborateur malveillant ou même une attaque par ransomware ne peut pas compromettre les sauvegardes, même s'il a pu accéder au système ONTAP lui-même. Associée à une planification Snapshot fréquente, cette solution offre une protection des données extrêmement puissante avec un RPO très faible.

> Les snapshots inviolables ne peuvent pas être hiérarchisés à l'aide du pool de fabric.

Pour plus d'informations sur les instantanés inviolables, consultez le "Documentation ONTAP."

> Dans la nouvelle plateforme ASA, répliquez les snapshots sur un cluster distant à l'aide de la règle *vault*, puis verrouillez la destination pour rendre les snapshots inviolables.

Réplication SnapMirror

Les snapshots peuvent également être répliqués sur un système distant. Cela inclut des snapshots inviolables, où la période de conservation est appliquée et appliquée sur le système distant. Il en résulte les mêmes avantages en matière de protection des données que les snapshots locaux, mais les données se trouvent sur une seconde baie de stockage. Cela permet de s'assurer que la destruction de la baie d'origine ne compromet pas les sauvegardes.

Un deuxième système ouvre également de nouvelles options pour la sécurité administrative. Par exemple, certains clients NetApp isolent les informations d'authentification pour les systèmes de stockage primaire et secondaire. Aucun utilisateur administratif n'a accès aux deux systèmes, ce qui signifie qu'un administrateur malveillant ne peut pas supprimer toutes les copies des données.

Pour plus d'informations sur SnapMirror, reportez-vous au "Documentation ONTAP."

Ordinateurs virtuels de stockage

Un système de stockage ONTAP nouvellement configuré est similaire à un serveur VMware ESX nouvellement provisionné, car aucun utilisateur ne peut prendre en charge avant la création d'une machine virtuelle. Avec ONTAP, vous créez une machine virtuelle de stockage (SVM) qui devient l'unité de gestion du stockage la plus élémentaire. Chaque SVM dispose de ses propres ressources de stockage, configurations de protocoles, adresses IP et WWN FCP. Il s'agit de la base de la colocation ONTAP.

Par exemple, vous pouvez configurer un SVM pour les charges de travail de production stratégiques et un second SVM sur un autre segment réseau pour les activités de développement. Vous pouvez alors restreindre l'accès au SVM de production à certains administrateurs, tout en accordant aux développeurs un contrôle plus étendu sur les ressources de stockage du SVM de développement. Vous devrez peut-être également proposer un troisième SVM à vos équipes financières et RH afin de stocker des données sensibles uniquement.

Pour plus d'informations sur les SVM, reportez-vous au "Documentation ONTAP."

RBAC d'administration

ONTAP offre un puissant contrôle d'accès basé sur des rôles (RBAC) pour les connexions d'administration. Certains administrateurs peuvent avoir besoin d'un accès complet au cluster, tandis que d'autres n'ont besoin que de l'accès à certains SVM. Le personnel du service d'assistance avancé peut avoir besoin d'augmenter la taille des volumes. Vous pouvez ainsi accorder aux utilisateurs administratifs l'accès requis pour s'acquitter de leurs responsabilités professionnelles, et rien de plus. De plus, vous pouvez sécuriser ces connexions à l'aide de PKI provenant de différents fournisseurs, restreindre l'accès aux clés ssh uniquement et appliquer les verrouillages de tentatives de connexion échouées.

Pour plus d'informations sur le contrôle d'accès administratif, reportez-vous au "Documentation ONTAP."

Authentification multifacteur (MFA)

ONTAP et certains autres produits NetApp prennent désormais en charge l'authentification multifacteur (MFA) selon plusieurs méthodes. Le résultat est un nom d'utilisateur/mot de passe compromis seul n'est pas un thread de sécurité sans les données du deuxième facteur, tel qu'un FOB ou une application de smartphone.

Pour plus d’informations sur l’authentification multifacteur (MFA), consultez le "Documentation ONTAP."

RBAC D’API

L’automatisation nécessite des appels d’API, mais tous les outils ne nécessitent pas un accès administratif complet. Pour sécuriser les systèmes d’automatisation, le RBAC est également disponible au niveau des API. Vous pouvez limiter les comptes d’utilisateur d’automatisation aux appels d’API requis. Par exemple, le logiciel de surveillance n’a pas besoin d’un accès de modification, il ne nécessite qu’un accès en lecture. Les flux de travail qui provisionnent le stockage n’ont pas besoin d’être supprimés.

Pour plus d’informations sur API RBAC, reportez-vous au "Documentation ONTAP."

Vérification multiadministrateur

L’authentification multifacteur peut être encore plus poussée en exigeant que deux administrateurs différents, chacun disposant de leurs propres informations d’identification, approuvent certaines activités. Cela inclut la modification des autorisations de connexion, l’exécution des commandes de diagnostic et la suppression des données.

Pour plus d’informations sur la vérification multiadministrateur (MAV), reportez-vous au "Documentation ONTAP."

MySQL

Présentation

MySQL et ses variantes, y compris MariaDB et Percona MySQL, est la base de données la plus populaire au monde.

> Cette documentation sur ONTAP et la base de données MySQL remplace la base de données *TR-4722: Base de données MySQL sur les meilleures pratiques ONTAP.*

ONTAP est une plate-forme idéale pour les bases de données MySQL car ONTAP est littéralement conçu pour les bases de données. De nombreuses fonctionnalités, telles que l'optimisation de la latence d'E/S aléatoire, pour la qualité de service (QoS) avancée et les fonctionnalités FlexClone de base, ont été spécialement conçues pour répondre aux besoins des charges de travail des bases de données.

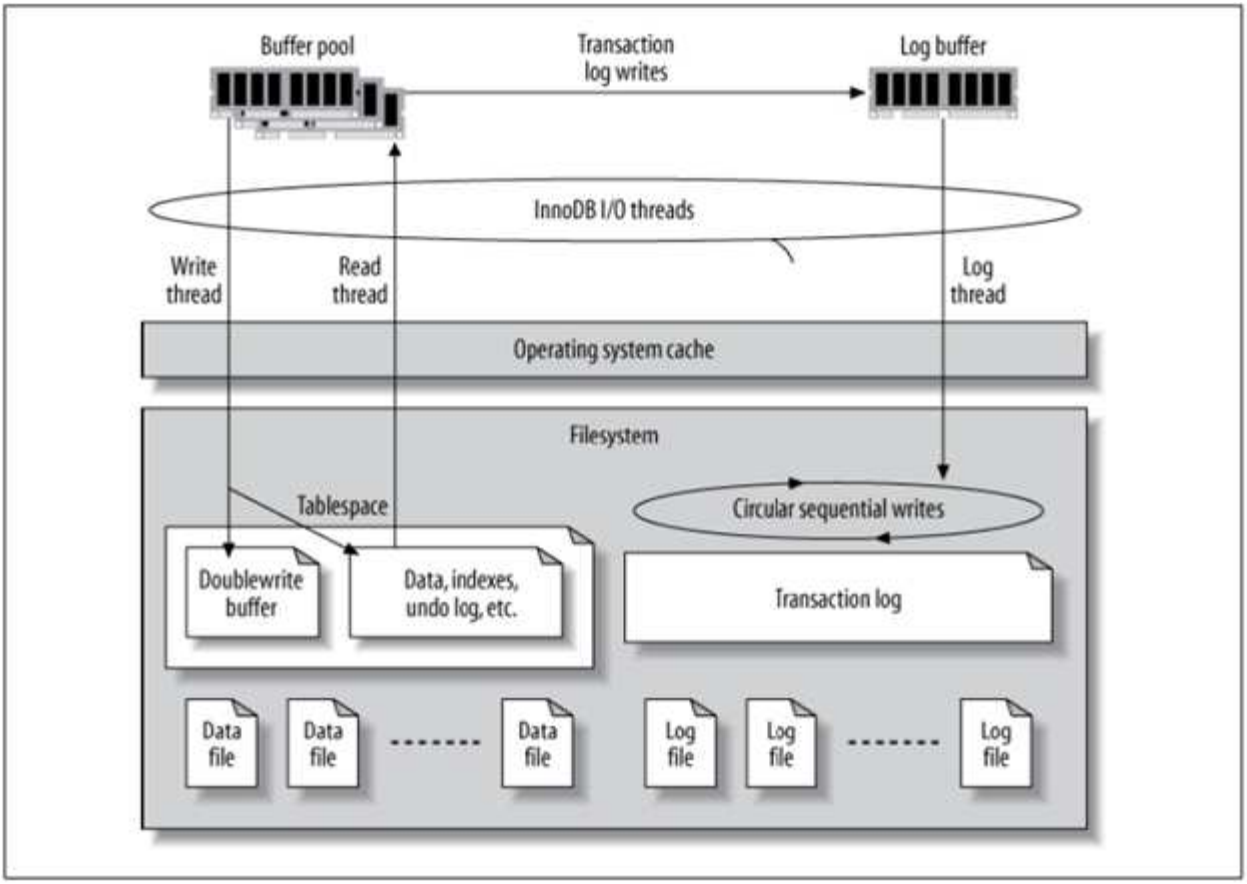
Des fonctionnalités supplémentaires, telles que les mises à niveau sans interruption (y compris le remplacement du stockage), assurent la disponibilité de vos bases de données stratégiques. Vous pouvez également bénéficier d'une reprise après incident instantanée pour les environnements volumineux via MetroCluster ou sélectionner des bases de données à l'aide de la synchronisation active SnapMirror.

Plus important encore, ONTAP offre des performances inégalées avec la possibilité de dimensionner la solution en fonction de vos besoins spécifiques. Nos systèmes haut de gamme peuvent fournir plus de 1 million d'IOPS à des latences mesurées en microsecondes. Toutefois, si vous n'avez besoin que de 100 000 IOPS, vous pouvez dimensionner correctement votre solution de stockage avec un contrôleur plus petit, qui exécute toujours le même système d'exploitation du stockage.

Configuration de la base de données

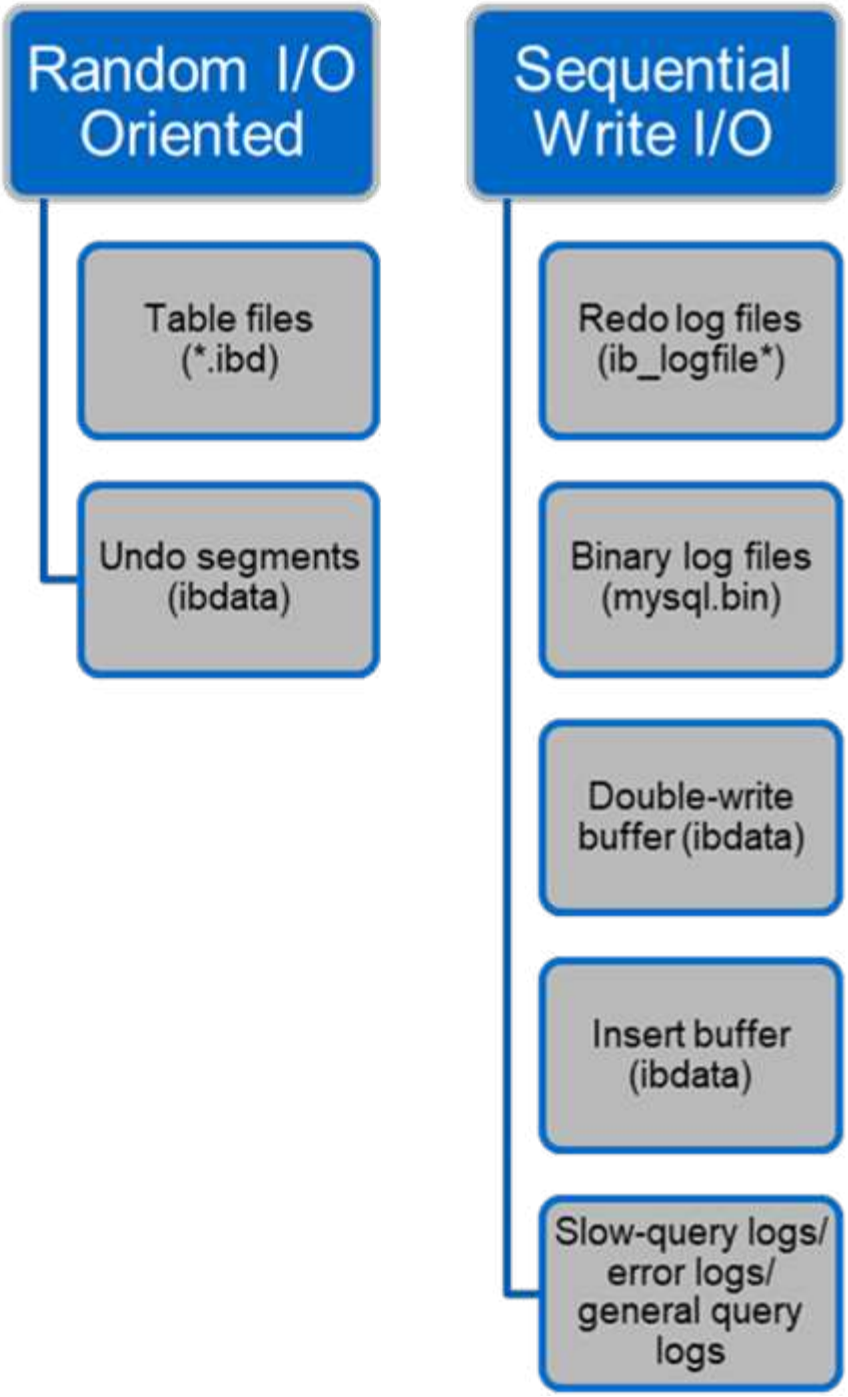
Structure de fichier

InnoDB agit comme la couche intermédiaire entre le stockage et le serveur MySQL, il stocke les données sur les lecteurs.



Les E/S MySQL sont classées en deux types :

- E/S de fichiers aléatoires
- E/S séquentielles de fichiers



Les fichiers de données sont lus et écrasés de manière aléatoire, ce qui entraîne un nombre élevé d'IOPS. Un stockage SSD est donc recommandé.

Les fichiers redo log et les fichiers log binaires sont des journaux transactionnels. Ils sont écrits de manière séquentielle, ce qui vous permet d'obtenir de bonnes performances sur le disque dur avec le cache d'écriture. Une lecture séquentielle a lieu lors de la restauration, mais cela provoque rarement un problème de performance, car la taille du fichier journal est généralement inférieure à celle des fichiers de données et les lectures séquentielles sont plus rapides que les lectures aléatoires (se produisant sur les fichiers de données).

La mémoire tampon en double écriture est une fonction spéciale d'InnoDB. InnoDB écrit d'abord les pages vidées dans le tampon de double écriture, puis écrit les pages à leur position correcte sur les fichiers de

données. Ce processus empêche la corruption de la page. Sans le tampon de double écriture, la page peut être corrompue si une panne de courant se produit pendant le processus d'écriture sur les lecteurs. L'écriture sur la mémoire tampon en double écriture étant séquentielle, elle est optimisée pour les disques durs. Les lectures séquentielles ont lieu lors de la restauration.

Comme la mémoire NVRAM ONTAP fournit déjà une protection en écriture, la mise en mémoire tampon en double écriture n'est pas nécessaire. MySQL a un paramètre, `skip_innodb_doublewrite`, pour désactiver le tampon de double écriture. Cette fonction peut améliorer considérablement les performances.

Le tampon d'insertion est également une fonction spéciale d'InnoDB. Si des blocs d'index secondaires non uniques ne sont pas en mémoire, InnoDB insère des entrées dans le tampon d'insertion pour éviter les opérations d'E/S aléatoires. Périodiquement, le tampon d'insertion est fusionné dans les arborescences d'index secondaires de la base de données. La mémoire tampon d'insertion réduit le nombre d'opérations d'E/S en fusionnant les demandes d'E/S vers le même bloc ; les opérations d'E/S aléatoires peuvent être séquentielles. Le tampon d'insertion est également hautement optimisé pour les disques durs. Les écritures et les lectures séquentielles ont lieu pendant les opérations normales.

Les segments d'annulation sont orientés E/S aléatoires. Pour garantir la simultanéité multiversion (MVCC), InnoDB doit enregistrer les anciennes images dans les segments d'annulation. La lecture des images précédentes à partir des segments d'annulation nécessite des lectures aléatoires. Si vous exécutez une longue transaction avec des lectures reproductibles (comme mysqldump—single transaction) ou exécutez une longue requête, les lectures aléatoires peuvent se produire. Par conséquent, le stockage des segments d'annulation sur des disques SSD est préférable dans ce cas. Si vous exécutez uniquement des transactions ou des requêtes courtes, les lectures aléatoires ne sont pas un problème.

NetApp recommande la disposition de conception de stockage suivante en raison des caractéristiques d'E/S InnoDB.

- Un volume pour stocker des fichiers MySQL orientés E/S aléatoires et séquentielles
- Un autre volume pour stocker des fichiers MySQL orientés E/S purement séquentiels

Cette disposition vous aide également à concevoir des stratégies et des règles de protection des données.

Paramètres de configuration

NetApp recommande quelques paramètres de configuration MySQL importants pour obtenir des performances optimales.

Paramètres	Valeurs
taille_fichier_log_innodb	256M
innodb_flush_log_at_trx_commit	2
innodb_doublewrite	0
innodb_flush_method	fsync
innodb_buffer_pool_size	11G
innodb_io_capacity	8192
innodb_buffer_pool_instances	8
innodb_lru_scan_depth	8192

open_file_limit	65535

Pour définir les paramètres décrits dans cette section, vous devez les modifier dans le fichier de configuration MySQL (my.cnf). Les meilleures pratiques NetApp sont le résultat de tests réalisés en interne.

taille_fichier_log_innodb

Il est important de sélectionner la bonne taille pour le fichier journal InnoDB pour les opérations d'écriture et pour avoir un temps de récupération décent après une panne du serveur.

Étant donné que tant de transactions sont connectées au fichier, la taille du fichier journal est importante pour les opérations d'écriture. Lorsque les enregistrements sont modifiés, la modification n'est pas immédiatement réécrite dans l'espace de table. Au lieu de cela, la modification est enregistrée à la fin du fichier journal et la page est marquée comme sale. InnoDB utilise son journal pour convertir les E/S aléatoires en E/S séquentielles

Lorsque le journal est plein, la page sale est écrite dans l'espace table en séquence pour libérer de l'espace dans le fichier journal. Par exemple, supposons qu'un serveur se bloque au milieu d'une transaction et que les opérations d'écriture ne sont enregistrées que dans le fichier journal. Avant que le serveur puisse de nouveau être mis en service, il doit passer par une phase de récupération dans laquelle les modifications enregistrées dans le fichier journal sont relus. Plus le nombre d'entrées dans le fichier journal est important, plus la restauration du serveur prend de temps.

Dans cet exemple, la taille du fichier journal affecte à la fois le temps de restauration et les performances d'écriture. Lorsque vous choisissez le bon nombre pour la taille du fichier journal, équilibrez le délai de restauration par rapport aux performances d'écriture. En général, tout ce qui se trouve entre 128M et 512M est d'une bonne valeur.

innodb_flush_log_at_trx_commit

En cas de modification des données, celles-ci ne sont pas immédiatement écrites sur le support de stockage.

À la place, les données sont enregistrées dans une mémoire tampon, qui est une partie de la mémoire allouée par InnoDB aux modifications de mémoire tampon enregistrées dans le fichier journal. InnoDB vide le tampon dans le fichier journal lorsqu'une transaction est validée, lorsque le tampon est plein ou une fois par seconde, quel que soit l'événement qui se produit en premier. La variable de configuration qui contrôle ce processus est innodb_flush_log_at_trx_commit. Les options de valeur comprennent :

- Lorsque vous réglez `innodb_flush_log_trx_at_commit=0`, InnoDB écrit les données modifiées (dans le pool de mémoire tampon InnoDB) dans le fichier journal (ib_logfile) et purge le fichier journal (écriture dans le stockage) toutes les secondes. Cependant, elle ne fait rien lorsque la transaction est validée. En cas de panne de courant ou de panne du système, aucune des données non rincées n'est récupérable car elles ne sont pas écrites sur le fichier journal ou les lecteurs.
- Lorsque vous réglez `innodb_flush_log_trx_commit=1`, InnoDB écrit la mémoire tampon du journal dans le journal de transactions et vide jusqu'à un stockage durable pour chaque transaction. Par exemple, pour toutes les validations de transactions, InnoDB écrit dans le journal, puis écrit dans le stockage. Un stockage plus lent affecte négativement les performances. Par exemple, le nombre de transactions InnoDB par seconde est réduit.
- Lorsque vous réglez `innodb_flush_log_trx_commit=2`, InnoDB écrit la mémoire tampon du journal

dans le fichier journal à chaque validation ; cependant, il n'écrit pas de données dans le stockage. InnoDB vide les données une fois par seconde. Même en cas de panne de courant ou de panne du système, les données de l'option 2 sont disponibles dans le fichier journal et peuvent être récupérées.

Si la performance est l'objectif principal, définissez la valeur sur 2. Comme InnoDB écrit sur les disques une fois par seconde, pas pour chaque validation de transaction, les performances s'améliorent considérablement. En cas de panne de courant ou de panne de courant, les données peuvent être récupérées à partir du journal de transactions.

Si la sécurité des données est l'objectif principal, définissez la valeur sur 1 afin que, pour chaque validation de transaction, InnoDB vide les lecteurs. Cependant, les performances peuvent être affectées.

NetApp recommande de définir la valeur innodb_flush_log_trx_commit sur 2 pour de meilleures performances.

innodb_doublewrite

Quand `innodb_doublewrite` Est activé (valeur par défaut), InnoDB stocke toutes les données deux fois : d'abord dans le tampon de double écriture, puis dans les fichiers de données réels.

Vous pouvez désactiver ce paramètre avec `--skip-innodb_doublewrite` pour les bancs d'essai ou lorsque vous êtes davantage préoccupé par les performances que par l'intégrité des données ou par les défaillances possibles. InnoDB utilise une technique de vidage de fichier appelée double écriture. Avant d'écrire des pages dans les fichiers de données, InnoDB les écrit dans une zone contiguë appelée tampon de double écriture. Une fois l'écriture et le vidage de la mémoire tampon en double écriture terminés, InnoDB écrit les pages dans leur position correcte dans le fichier de données. Si le système d'exploitation ou un processus mysqld se bloque lors d'une écriture de page, InnoDB peut plus tard trouver une bonne copie de la page à partir du tampon de double écriture pendant la récupération après panne.

NetApp recommande de désactiver le tampon en double écriture. La mémoire NVRAM de ONTAP remplit la même fonction. La double mise en mémoire tampon endommagera inutilement les performances.

innodb_buffer_pool_size

Le pool de mémoire tampon InnoDB est la partie la plus importante de toute activité de réglage.

InnoDB s'appuie fortement sur le pool de mémoire tampon pour mettre en cache les index et ramer les données, l'index de hachage adaptatif, le tampon d'insertion et de nombreuses autres structures de données utilisées en interne. Le pool de mémoire tampon met également en mémoire tampon les modifications apportées aux données afin que les opérations d'écriture n'aient pas à être exécutées immédiatement sur le stockage, ce qui améliore les performances. Le pool de mémoire tampon fait partie intégrante d'InnoDB et sa taille doit être ajustée en conséquence. Tenez compte des facteurs suivants lors de la définition de la taille du pool de mémoire tampon :

- Pour une machine dédiée uniquement InnoDB, définissez la taille du pool de mémoire tampon sur 80 % ou plus de la mémoire RAM disponible.
- S'il ne s'agit pas d'un serveur dédié MySQL, définissez la taille sur 50 % de RAM.

innodb_flush_method

Le paramètre innodb_flush_method indique comment InnoDB ouvre et vide les fichiers journaux et de données.

Optimisations

Dans l'optimisation InnoDB, la définition de ce paramètre permet de régler les performances de la base de données, le cas échéant.

Les options suivantes permettent de vider les fichiers via InnoDB :

- `fsync`. InnoDB utilise le `fsync()` appel système pour vider les fichiers de données et les fichiers journaux. Cette option est le paramètre par défaut.
- `O_DSYNC`. InnoDB utilise le `O_DSYNC` option permettant d'ouvrir et de vider les fichiers journaux et fsync() pour vider les fichiers de données. InnoDB n'utilise pas `O_DSYNC` Directement, parce qu'il y a eu des problèmes avec elle sur de nombreuses variétés d'UNIX.
- `O_DIRECT`. InnoDB utilise le `O_DIRECT` option (ou `directio()` Sous Solaris) pour ouvrir les fichiers de données et les utilise `fsync()` pour vider les fichiers de données et les fichiers journaux. Cette option est disponible sur certaines versions de GNU/Linux, FreeBSD et Solaris.
- `O_DIRECT_NO_FSYNC`. InnoDB utilise le `O_DIRECT` Option pendant le vidage des E/S ; cependant, il ignore le `fsync()` appel système par la suite. Cette option n'est pas adaptée à certains types de systèmes de fichiers (par exemple, XFS). Si vous n'êtes pas sûr que votre système de fichiers nécessite un `fsync()` l'appel système, par exemple pour conserver toutes les métadonnées de fichier, utilise le `O_DIRECT` à la place.

Observation

Dans les tests de laboratoire NetApp, le `fsync` L'option par défaut a été utilisée sur NFS et SAN, et il s'agissait d'un outil d'amélioration des performances par rapport à `O_DIRECT`. Lors de l'utilisation de la méthode de rinçage comme `O_DIRECT` Avec ONTAP, nous avons observé que le client écrit beaucoup d'écritures sur un seul octet à la frontière du bloc 4096 en série. Ces écritures ont augmenté la latence sur le réseau et dégradé les performances.

innodb_io_capacity

Dans le plug-in InnoDB, un nouveau paramètre appelé innodb_io_Capacity a été ajouté à partir de MySQL 5.7.

Il contrôle le nombre maximal d'IOPS qu'InnoDB exécute (qui inclut la vitesse de vidage des pages sales ainsi que la taille de lot du tampon d'insertion [ibuf]). Le paramètre innodb_io_Capacity définit une limite supérieure sur les IOPS par les tâches d'arrière-plan InnoDB, telles que le vidage des pages du pool de mémoire tampon et la fusion des données à partir du tampon de changement.

Définissez le paramètre innodb_io_Capacity sur le nombre approximatif d'opérations d'E/S que le système peut effectuer par seconde. Idéalement, maintenez le paramètre aussi bas que possible, mais pas si bas que les activités en arrière-plan ralentissent. Si le paramètre est trop élevé, les données sont supprimées du pool de mémoire tampon et la mémoire tampon est insérée trop rapidement pour que la mise en cache offre un avantage significatif.

NetApp recommande que si vous utilisez ce paramètre sur NFS, analysez le résultat du test d'IOPS (SysBench/FiO) et définissez le paramètre en conséquence. Utilisez la plus petite valeur possible pour le vidage et la purge pour continuer à fonctionner, sauf si vous voyez plus de pages modifiées ou sales que vous le souhaitez dans le pool de mémoire tampon InnoDB.

N'utilisez pas de valeurs extrêmes telles que 20,000 ou plus, sauf si vous avez prouvé que des valeurs inférieures ne suffisent pas à votre charge de travail.

Le paramètre InnoDB_IO_Capacity régule les débits de rinçage et les E/S associées

Vous pouvez sérieusement nuire aux performances en définissant ce paramètre ou le paramètre innodb_io_Capacity_max trop élevé et gaspillé

innodb_lru_scan_depth

Le `innodb_lru_scan_depth` Le paramètre influence les algorithmes et les heuristiques de l'opération de vidage pour le pool de mémoire tampon InnoDB.

Ce paramètre intéresse principalement les experts en performances qui s'intéressent au réglage des charges de travail exigeantes en E/S. Pour chaque instance de pool de mémoire tampon, ce paramètre indique la distance vers le bas dans la liste de pages LRU (least recently used) que le thread de nettoyage de page doit poursuivre la numérisation, en recherchant les pages sales à vider. Cette opération d'arrière-plan est effectuée une fois par seconde.

Vous pouvez régler la valeur vers le haut ou vers le bas pour réduire le nombre de pages libres. Ne définissez pas la valeur beaucoup plus haut que nécessaire, car les analyses peuvent avoir un coût de performance important. Pensez également à ajuster ce paramètre lors de la modification du nombre d'instances de pool de mémoire tampon, car `innodb_lru_scan_depth * innodb_buffer_pool_instances` définit la quantité de travail effectuée par le thread de nettoyage de page chaque seconde.

Un paramètre inférieur à celui par défaut convient à la plupart des workloads. Envisagez d'augmenter la valeur uniquement si vous disposez d'une capacité d'E/S disponible pour une charge de travail classique. Inversement, si une charge de travail exigeante en écriture sature votre capacité d'E/S, diminuez la valeur, en particulier si vous disposez d'un pool de mémoire tampon important.

open_file_limits

Le `open_file_limits` paramètre détermine le nombre de fichiers que le système d'exploitation autorise à ouvrir mysqld.

La valeur de ce paramètre au moment de l'exécution est la valeur réelle autorisée par le système et peut être différente de la valeur spécifiée au démarrage du serveur. La valeur est 0 sur les systèmes où MySQL ne peut pas modifier le nombre de fichiers ouverts. L'efficace `open_files_limit` la valeur est basée sur la valeur spécifiée au démarrage du système (le cas échéant) et sur les valeurs de `max_connections` et `table_open_cache` en utilisant ces formules :

- 10 + `max_connections` + (`table_open_cache` x 2)
- `max_connections` x 5
- Limite du système d'exploitation si positif

- Si la limite du système d'exploitation est infinie : `open_files_limit` la valeur est spécifiée au démarrage ; 5,000 si aucune

Le serveur tente d'obtenir le nombre de descripteurs de fichier en utilisant le maximum de ces quatre valeurs. Si ce nombre de descripteurs ne peut pas être obtenu, le serveur tente d'obtenir autant que le système le permet.

Configuration de l'hôte

Conteneurisation

La conteneurisation des bases de données MySQL est de plus en plus répandue.

La gestion des conteneurs de faible niveau est presque toujours effectuée via Docker. Les plateformes de gestion de conteneurs comme OpenShift et Kubernetes simplifient encore la gestion des grands environnements de conteneurs. La conteneurisation présente des avantages à moindre coût, car il n'est pas nécessaire de posséder une licence pour un hyperviseur. De plus, les conteneurs permettent à plusieurs bases de données de s'exécuter isolées les unes des autres tout en partageant le même noyau et le même système d'exploitation sous-jacents. Les conteneurs sont provisionnés en quelques microsecondes.

NetApp propose Astra Trident pour fournir des fonctionnalités de gestion avancées du stockage. Par exemple, Astra Trident permet à un conteneur créé dans Kubernetes de provisionner automatiquement son stockage sur le Tier approprié, d'appliquer des règles d'exportation, de définir des règles Snapshot et même de cloner un conteneur vers un autre. Pour plus d'informations, reportez-vous à la "Documentation Astra Trident".

Tables d'emplacements NFSv3

Les performances de NFSv3 sous Linux dépendent d'un paramètre appelé `tcp_max_slot_table_entries`.

Les tables d'emplacements TCP sont l'équivalent NFSv3 de la profondeur de file d'attente de l'adaptateur de bus hôte (HBA). Ces tableaux contrôlent le nombre d'opérations NFS qui peuvent être en attente à la fois. La valeur par défaut est généralement 16, un chiffre bien trop faible pour assurer des performances optimales. Le problème inverse se produit sur les noyaux Linux plus récents : la limite de la table des emplacements TCP augmente automatiquement par envoi de demandes, jusqu'à atteindre le niveau de saturation du serveur NFS.

Pour des performances optimales et pour éviter les problèmes de performances, ajustez les paramètres du noyau qui contrôlent les tables d'emplacements TCP.

Exécutez le `sysctl -a | grep tcp.*.slot_table` et observez les paramètres suivants :

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tous les systèmes Linux doivent inclure `sunrpc.tcp_slot_table_entries`, mais seulement certains incluent `sunrpc.tcp_max_slot_table_entries`. Ils doivent tous deux être réglés sur 128.

Si vous ne définissez pas ces paramètres, vous risquez d'avoir des effets importants sur les performances. Dans certains cas, les performances sont limitées car le système d'exploitation linux n'émet pas suffisamment d'E/S. Dans d'autres cas, les latences d'E/S augmentent à mesure que le système d'exploitation linux tente d'émettre plus d'E/S que ce qui peut être traité.

Planificateurs d'E/S.

Le noyau Linux permet un contrôle de bas niveau sur la façon dont les E/S sont planifiées pour bloquer les périphériques.

Les valeurs par défaut sur les différentes distributions de Linux varient considérablement. MySQL vous recommande d'utiliser `NOOP` ou un `deadline` Planificateur d'E/S avec E/S asynchrones natives (AIO) sous Linux. De manière générale, les clients NetApp et les tests internes montrent de meilleurs résultats avec NoOps.

Le moteur de stockage InnoDB de MySQL utilise le sous-système d'E/S asynchrone (AIO natif) sur Linux pour effectuer des demandes de lecture et d'écriture pour les pages de fichiers de données. Ce comportement est contrôlé par le `innodb_use_native_aio` option de configuration, activée par défaut. Avec le tout-en-un natif, le type de planificateur d'E/S a une plus grande influence sur les performances E/S. Menez des bancs d'essai pour déterminer quel planificateur d'E/S offre les meilleurs résultats pour votre charge de travail et votre environnement.

Consultez la documentation Linux et MySQL appropriée pour obtenir des instructions sur la configuration du planificateur d'E/S.

Descripteurs de fichier

Pour s'exécuter, le serveur MySQL a besoin de descripteurs de fichier et les valeurs par défaut ne sont pas suffisantes.

Il les utilise pour ouvrir de nouvelles connexions, stocker des tables dans le cache, créer des tables temporaires pour résoudre des requêtes complexes et accéder à des requêtes persistantes. Si mysqld n'est pas en mesure d'ouvrir de nouveaux fichiers lorsque cela est nécessaire, il peut arrêter de fonctionner correctement. Un symptôme courant de ce problème est l'erreur 24, "trop de fichiers ouverts". Le nombre de descripteurs de fichier que mysqld peut ouvrir simultanément est défini par le `open_files_limit` option définie dans le fichier de configuration (`/etc/my.cnf`). Mais `open_files_limit` dépend également des limites du système d'exploitation. Cette dépendance complique la définition de la variable.

MySQL ne peut pas définir son `open_files_limit` option supérieure à celle spécifiée sous `ulimit 'open files'`. Par conséquent, vous devez définir explicitement ces limites au niveau du système d'exploitation pour permettre à MySQL d'ouvrir des fichiers si nécessaire. Il existe deux façons de vérifier la limite de fichiers sous Linux :

- Le `ulimit` commande vous donne rapidement une description détaillée des paramètres autorisés ou verrouillés. Les modifications apportées par l'exécution de cette commande ne sont pas permanentes et seront effacées après un redémarrage du système.
- Modifications apportées au `/etc/security/limit.conf` les fichiers sont permanents et ne sont pas affectés par un redémarrage du système.

Assurez-vous de modifier les limites matérielles et logicielles de l'utilisateur mysql. Les extraits suivants sont issus de la configuration :

```
mysql hard nofile 65535
mysql soft nofile 65353
```

En parallèle, mettez à jour la même configuration dans `my.cnf` pour utiliser pleinement les limites de fichiers ouverts.

Configuration de stockage sous-jacente

NFS

La documentation MySQL recommande d'utiliser NFSv4 pour les déploiements NAS.

Tailles de transfert NFS ONTAP

Par défaut, ONTAP limite les tailles d'E/S NFS à 64 Ko. Les E/S aléatoires avec une base de données MySQL utilisent une taille de bloc bien inférieure à la taille maximale de 64 Ko. Les E/S de bloc volumineux sont généralement parallélisées de sorte que le maximum de 64 000 ne constitue pas non plus une limitation.

Dans certains cas, le maximum de 64 000 charges de travail entraîne une limitation. En particulier, les opérations à thread unique, telles que les opérations de sauvegarde d'analyse de table complète, s'exécuteront plus rapidement et plus efficacement si la base de données peut exécuter moins d'E/S, mais de plus grande taille. La taille optimale de gestion des E/S pour ONTAP avec charges de travail de base de données est de 256 Ko. Les options de montage NFS répertoriées pour les systèmes d'exploitation spécifiques ci-dessous ont été mises à jour de 64 Ko à 256 Ko en conséquence.

La taille maximale de transfert pour un SVM ONTAP donné peut être modifiée comme suit :

```
Cluster01::> set advanced

Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.

Do you want to continue? {y|n}: y

Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
```

> Ne diminuez jamais la taille de transfert maximale autorisée sur ONTAP en dessous de la valeur de rsize/wsize des systèmes de fichiers NFS actuellement montés. Cela peut provoquer des blocages ou même une corruption des données avec certains systèmes d'exploitation. Par exemple, si les clients NFS sont actuellement définis sur une taille rsize/wsize de 65536, la taille maximale du transfert ONTAP peut être ajustée entre 65536 et 1048576 sans effet car les clients eux-mêmes sont limités. Réduire la taille de transfert maximale en dessous de 65536 peut endommager la disponibilité ou les données.

NetApp recommande

Définition du paramètre NFSv4 fstab (/etc/fstab) suivant :

```
nfs4 rw,
hard,nointr,bg,vers=4,proto=tcp,noatime,rsize=262144,wsize=262144
```

NFSv3 présentait fréquemment un problème de verrouillage des fichiers journaux InnoDB après une panne de courant. L'utilisation du temps ou la commutation des fichiers journaux a résolu ce problème. Cependant, NFSv4 dispose d'opérations de verrouillage et assure le suivi des fichiers ouverts et des délégations.

SAN

Les bases de données plus petites peuvent être placées sur une paire de LUN standard tant que les besoins en E/S et en capacité se situent dans les limites d'un seul système de fichiers LUN. Par exemple, une base de données qui nécessite environ 2 000 IOPS aléatoires peut être hébergée sur un système de fichiers unique sur une seule LUN. De même, une base de données de 100 Go seulement serait prise en charge sur une seule LUN sans problème de gestion.

Les bases de données plus grandes nécessitent plusieurs LUN. Par exemple, une base de données qui nécessite 100 000 IOPS aurait probablement besoin d'au moins huit LUN. Un seul LUN deviendrait un goulot d'étranglement en raison du nombre insuffisant de canaux SCSI vers les disques. Une base de données de 10 To serait tout aussi difficile à gérer sur un seul LUN de 10 To. Les gestionnaires de volumes logiques sont conçus pour lier les performances et les capacités de plusieurs LUN afin d'améliorer les performances et la gestion.

Dans les deux cas, une paire de volumes ONTAP doit suffire. Dans une configuration simple, le LUN du fichier de données serait placé dans un volume dédié, tout comme le LUN du journal. Avec une configuration de gestionnaire de volumes logique, toutes les LUN du groupe de volumes de fichiers de données se trouvent dans un volume dédié, et les LUN du groupe de volumes de logs se trouvent dans un second volume dédié.

NetApp recommande l'utilisation de deux systèmes de fichiers pour les déploiements MySQL sur SAN :

- Le premier système de fichiers stocke toutes les données MySQL, y compris l'espace table, les données et l'index.
- Le second système de fichiers stocke tous les journaux (journaux binaires, journaux lents et journaux des transactions).

Il existe plusieurs raisons de séparer les données de cette manière :

- Les modèles d'E/S des fichiers de données et des fichiers journaux diffèrent. De les séparer, on disposerait d'un plus grand nombre d'options avec les contrôles de QoS.
- Pour optimiser l'utilisation de la technologie Snapshot, vous devez pouvoir restaurer les fichiers de données de manière indépendante. La connexion de fichiers de données avec des fichiers journaux interfère avec la restauration des fichiers de données.
- La technologie SnapMirror de NetApp peut être utilisée pour fournir une fonctionnalité de reprise d'activité simple à faible RPO pour une base de données. Toutefois, le planning de réplication des fichiers de données et des journaux doit être différent.

Utilisez cette disposition de base à deux volumes pour pérenniser votre solution et utiliser toutes les fonctionnalités ONTAP si nécessaire.

NetApp recommande de formater votre lecteur avec le système de fichiers ext4 en raison des fonctionnalités suivantes :

- Approche étendue des fonctions de gestion des blocs utilisées dans le système de fichiers de journalisation (JFS) et fonctions d'allocation différée du système de fichiers étendu (XFS).
- Ext4 autorise des systèmes de fichiers allant jusqu'à 1 exbioctet (2^60 octets) et des fichiers allant jusqu'à 16 tébioctets (16 * 2^40 octets). En revanche, le système de fichiers ext3 ne prend en charge qu'une taille de système de fichiers maximale de 16 To et une taille de fichier maximale de 2 To.
- Dans les systèmes de fichiers ext4, l'allocation multi-blocs (mballoc) alloue plusieurs blocs pour un fichier en une seule opération, au lieu de les allouer un par un, comme dans ext3. Cette configuration réduit la surcharge liée à l'appel de l'ALlocator de bloc plusieurs fois et optimise l'allocation de mémoire.
- Bien que XFS soit la valeur par défaut pour de nombreuses distributions Linux, il gère les métadonnées différemment et ne convient pas à certaines configurations MySQL.

NetApp recommande d'utiliser les options de taille de bloc de 4 ko avec l'utilitaire mkfs pour l'aligner avec la taille de LUN de bloc existante.

```
mkfs.ext4 -b 4096
```

Les LUN NetApp stockent les données dans des blocs physiques de 4 Ko, ce qui produit huit blocs logiques de 512 octets.

Si vous ne configurez pas la même taille de bloc, les E/S ne seront pas alignées avec les blocs physiques correctement et pourraient écrire sur deux disques différents dans un groupe RAID, ce qui entraîne une latence.

Il est important d'aligner les E/S pour que les opérations de lecture/écriture soient fluides. Cependant, lorsque les E/S commencent au niveau d'un bloc logique qui n'est pas au début d'un bloc physique, les E/S sont mal alignées. Les opérations d'E/S ne sont alignées que lorsqu'elles commencent au niveau d'un bloc logique, le premier bloc logique d'un bloc physique.

Base de données Oracle

Bases de données Oracle sur ONTAP

ONTAP est conçu pour les bases de données Oracle. Pendant des décennies, ONTAP a été optimisé pour les demandes uniques d'E/S de bases de données relationnelles. Plusieurs fonctionnalités ONTAP ont été créées spécifiquement pour répondre aux besoins des bases de données Oracle, et même à la demande d'Oracle Inc. Elle-même.

> Cette documentation remplace les rapports techniques publiés précédemment *TR-3633 : bases de données Oracle sur ONTAP ; TR-4591 : protection des données Oracle : sauvegarde, restauration, réplication ; TR-4592 : Oracle sur MetroCluster ; et TR-4534 : migration de bases de données Oracle vers des systèmes de stockage NetApp*

Outre les nombreuses possibilités offertes par ONTAP pour valoriser votre environnement de base de données, les besoins des utilisateurs sont très variés, notamment en termes de taille de la base de données, de performances et de protection des données. Les déploiements de systèmes de stockage NetApp prennent des formes diverses, qu'il s'agisse d'un environnement virtualisé incluant environ 6,000 bases de données fonctionnant sous VMware ESX ou d'un data warehouse à instance unique dont la taille de 996 To ne cesse de croître. Par conséquent, il existe peu de bonnes pratiques claires pour la configuration d'une base de données Oracle sur un système de stockage NetApp.

Les exigences relatives à l'exploitation d'une base de données Oracle sur un stockage NetApp sont traitées de deux manières. Tout d'abord, lorsqu'il existe une bonne pratique claire, elle sera appelée spécifiquement. D'une manière générale, de nombreuses considérations de conception à prendre en compte par les architectes de solutions de stockage Oracle en fonction de leurs besoins spécifiques seront expliquées.

Configuration ONTAP

RAID

RAID désigne l'utilisation de la redondance pour protéger les données contre la perte d'un disque.

Des questions se posent parfois au sujet des niveaux RAID dans la configuration du stockage NetApp utilisé pour les bases de données Oracle et d'autres applications d'entreprise. De nombreuses meilleures pratiques Oracle en matière de configuration de baie de stockage contiennent des avertissements concernant l'utilisation de la mise en miroir RAID et/ou l'évitement de certains types de RAID. Bien qu'elles soulèvent des points valides, ces sources ne s'appliquent pas au RAID 4 et aux technologies NetApp RAID DP et RAID-TEC utilisées dans ONTAP.

RAID 4, RAID 5, RAID 6, RAID DP et RAID-TEC utilisent tous la parité pour s'assurer qu'une panne de disque n'entraîne pas de perte de données. Ces options RAID offrent une meilleure utilisation du stockage que la mise en miroir, mais la plupart des implémentations RAID présentent des inconvénients pour les opérations d'écriture. La réalisation d'une opération d'écriture sur d'autres implémentations RAID peut nécessiter plusieurs lectures de disque pour régénérer les données de parité, un processus communément appelé la pénalité RAID.

Cependant, ONTAP n'entraîne pas cette pénalité RAID. Cela est dû à l'intégration de NetApp WAFL (Write Anywhere File Layout) à la couche RAID. Les opérations d'écriture sont fusionnées dans la mémoire RAM et

préparées sous la forme d'une couche RAID complète, y compris la génération de la parité. ONTAP n'a pas besoin d'effectuer de lecture pour effectuer une écriture, ce qui signifie que ONTAP et WAFL évitent la pénalité RAID. Les performances des opérations stratégiques pour la latence, telles que la journalisation de reprise, sont assurées sans aucun obstacle. Les écritures aléatoires des fichiers de données n'entraînent aucune pénalité RAID résultant de la régénération de la parité.

En ce qui concerne la fiabilité statistique, même RAID DP offre une meilleure protection que la mise en miroir RAID. Le problème principal est la demande sur disques lors de la reconstruction RAID. Avec une configuration RAID en miroir, le risque de perte de données en cas de défaillance d'un disque pendant la reconstruction vers son partenaire dans la configuration RAID est bien plus grand que le risque de défaillance simultanée de trois disques dans une configuration RAID DP.

Gestion de la capacité

La gestion d'une base de données ou d'une autre application d'entreprise avec un stockage d'entreprise prévisible, gérable et haute performance requiert de l'espace libre sur les disques pour la gestion des données et des métadonnées. La quantité d'espace libre requise dépend du type de disque utilisé et des processus métier.

L'espace libre est défini comme tout espace qui n'est pas utilisé pour les données réelles et inclut l'espace non alloué sur l'agrégat lui-même et l'espace inutilisé au sein des volumes constitutifs. Le provisionnement fin doit également être envisagé. Par exemple, un volume peut contenir une LUN de 1 To, dont seulement 50 % sont utilisés par des données réelles. Dans un environnement à provisionnement fin, cet espace semble être consommé de 500 Go. Toutefois, dans un environnement entièrement provisionné, la capacité totale de 1 To semble être utilisée. Les 500 Go d'espace non alloué sont masqués. Cet espace n'est pas utilisé par les données réelles et doit donc être inclus dans le calcul de l'espace libre total.

Les recommandations de NetApp pour les systèmes de stockage utilisés pour les applications d'entreprise sont les suivantes :

Des agrégats SSD, y compris les systèmes AFF

> **NetApp recommande** un minimum de 10% d'espace libre. Cela inclut tout l'espace inutilisé, y compris l'espace libre au sein de l'agrégat ou d'un volume, ainsi que tout espace libre alloué en raison de l'utilisation du provisionnement complet, mais qui n'est pas utilisé par les données réelles. L'espace logique n'est pas important, la question est de savoir quelle quantité d'espace physique réellement disponible pour le stockage des données.

La recommandation de 10 % d'espace libre est très prudente. Les agrégats SSD peuvent prendre en charge des charges de travail à des niveaux d'utilisation encore plus élevés, sans affecter les performances. Cependant, à mesure que l'utilisation de l'agrégat augmente, le risque de manquer d'espace augmente également si l'utilisation n'est pas surveillée de près. De plus, même si vous utilisez un système à 99 % de capacité, les performances risquent d'être moins élevées, mais vous devrez probablement interrompre la gestion pour l'empêcher de se remplir complètement lors de la commande de matériel supplémentaire. L'acquisition et l'installation de disques supplémentaires peuvent prendre un certain temps.

Les agrégats HDD, y compris les agrégats Flash Pool

NetApp recommande un minimum de 15 % d'espace libre lorsque des disques rotatifs sont utilisés. Cela inclut tout l'espace inutilisé, y compris l'espace libre au sein de l'agrégat ou d'un volume, ainsi que tout espace libre alloué en raison de l'utilisation du provisionnement complet, mais qui n'est pas utilisé par les données réelles. Les performances seront affectées à mesure que l'espace libre approche 10 %.

Ordinateurs virtuels de stockage

La gestion du stockage des bases de données Oracle est centralisée sur un SVM (Storage Virtual machine)

Un SVM, connu sous le nom de vserver sur l'interface de ligne de commandes ONTAP, est une unité fonctionnelle de base du stockage. Il est utile de comparer un SVM à un invité sur un serveur VMware ESX.

Lors de l'installation initiale, ESX ne possède pas de fonctionnalités préconfigurées, telles que l'hébergement d'un système d'exploitation invité ou la prise en charge d'une application utilisateur. Il s'agit d'un conteneur vide jusqu'à ce qu'une machine virtuelle (VM) soit définie. ONTAP fonctionne de manière similaire : Lors de la première installation de ONTAP, aucune fonctionnalité de service des données n'est disponible tant qu'un SVM n'est pas créé. Pour configurer les services de données.

À l'instar des autres aspects de l'architecture de stockage, les meilleures options pour la conception des SVM et de l'interface logique (LIF) dépendent largement des exigences d'évolutivité et des besoins de l'entreprise.

SVM

Il n'existe aucune bonne pratique officielle de provisionnement des SVM pour ONTAP. La bonne approche dépend des exigences en matière de gestion et de sécurité.

La plupart des clients utilisent un SVM principal pour la plupart de leurs besoins quotidiens, mais ils en créent un petit pour des besoins particuliers. Par exemple, vous pouvez créer :

- SVM d'une base de données stratégique gérée par une équipe de spécialistes
- SVM pour un groupe de développement auquel un contrôle administratif complet a été attribué afin de pouvoir gérer leur propre stockage indépendamment
- SVM pour les données sensibles de l'entreprise, telles que les données de rapports financiers ou de ressources humaines, pour lesquelles l'équipe administrative doit être limitée

Dans un environnement de colocation, on peut attribuer à chaque locataire une SVM dédiée aux données. La limite du nombre de SVM et de LIF par cluster, paire HA et nœud dépend du protocole utilisé, du modèle de nœud et de la version de ONTAP. Consulter le "NetApp Hardware Universe" pour ces limites.

Gestion des performances grâce à la QoS de ONTAP

Pour gérer efficacement et en toute sécurité plusieurs bases de données Oracle, il est nécessaire de disposer d'une stratégie de qualité de service efficace. C'est pourquoi les systèmes de stockage modernes offrent des performances toujours plus élevées.

Plus précisément, l'adoption croissante des systèmes de stockage 100 % Flash a permis de consolider les charges de travail. Les baies de stockage qui reposent sur des supports rotatifs ne prennent généralement en charge qu'un nombre limité de charges de travail exigeantes en E/S, car leurs capacités IOPS sont limitées par rapport aux anciens disques rotatifs. Une ou deux bases de données fortement actives saturaient les disques sous-jacents bien avant que les contrôleurs de stockage n'atteignent leurs limites. Cela a changé. Il

est possible de saturer les contrôleurs de stockage les plus puissants, car le nombre de disques SSD requis est relativement faible. Cela signifie que vous pouvez exploiter pleinement les capacités des contrôleurs sans craindre un effondrement soudain des performances lors de pics de latence des supports rotatifs.

À titre d'exemple de référence, un simple système AFF A800 HA à deux nœuds est capable de traiter jusqu'à un million d'IOPS aléatoires avant que la latence ne dépasse la milliseconde. On pourrait s'attendre à ce que très peu de charges de travail atteignent de tels niveaux. L'utilisation optimale de cette baie AFF A800 implique l'hébergement de plusieurs workloads. Pour ce faire, la sécurité et la prévisibilité exigent des contrôles de QoS.

Il existe deux types de qualité de service (QoS) dans ONTAP : les IOPS et la bande passante. Les contrôles de QoS peuvent être appliqués aux SVM, volumes, LUN et fichiers.

QoS des IOPS

Un contrôle de la QoS pour les IOPS est évidemment basé sur l'ensemble des IOPS d'une ressource donnée, mais il existe un certain nombre d'aspects de la QoS pour les IOPS qui peuvent ne pas être intuitifs. Au départ, quelques clients ont été surpris par l'augmentation apparente de la latence lorsqu'un seuil d'IOPS est atteint. L'augmentation de la latence est la conséquence naturelle de la limitation des IOPS. Logiquement, il fonctionne de la même manière qu'un système de jetons. Par exemple, si un volume donné contenant des fichiers de données dispose d'une limite de 10 000 IOPS, chaque E/S arrivant doit d'abord recevoir un jeton pour poursuivre le traitement. Tant que plus de 10 000 jetons n'ont pas été consommés en une seconde donnée, aucun retard n'est présent. Si les opérations d'E/S doivent attendre la réception de leur jeton, cet attente apparaît comme une latence supplémentaire. Plus une charge de travail est élevée, plus les E/S sont longues à attendre dans la file d'attente pour le traitement de son tour, ce qui apparaît comme une latence plus élevée.

> Soyez prudent lorsque vous appliquez des contrôles QoS aux données des transactions de base de données/journaux de reprise. Alors que les demandes de performances liées à la journalisation de reprise sont généralement très élevées, bien inférieures à celles des fichiers de données, l'activité du journal de reprise est en rafales. L'E/S se produit en de brèves impulsions et une limite de QoS qui semble appropriée pour les niveaux d'E/S de reprise moyens peut être trop basse pour les exigences réelles. Cela peut entraîner de strictes limitations de performance en cas d'engagement de la QoS avec chaque pic de journal de reprise. En général, la journalisation des opérations de reprise et d'archivage ne doit pas être limitée par la QoS.

QoS de la bande passante

Toutes les tailles d'E/S ne sont pas identiques. Par exemple, une base de données peut effectuer de nombreuses lectures de blocs de petite taille, ce qui entraînerait l'atteinte du seuil d'IOPS, mais il est également possible que les bases de données effectuent une analyse de table complète comprenant un très petit nombre de lectures de blocs volumineux, qui consomment une très grande quantité de bande passante, mais relativement peu d'IOPS.

De même, un environnement VMware peut générer un nombre très élevé d'IOPS aléatoires au démarrage, mais exécuter moins d'E/S, mais plus importantes, lors d'une sauvegarde externe.

Pour gérer efficacement les performances, les IOPS ou la bande passante doivent parfois être limitées, voire les deux.

QoS minimale/garantie

De nombreux clients recherchent une solution incluant une QoS garantie, qui semble plus difficile à atteindre qu'elle ne le paraît et qui risque d'être très gaspillée. Par exemple, pour placer 10 bases de données avec une

garantie de 10 000 IOPS, il est nécessaire de dimensionner un système dans le cas où les 10 bases de données s'exécutent simultanément à 10 000 IOPS, pour un total de 100 000.

La meilleure utilisation pour les contrôles QoS minimaux est de protéger les charges de travail stratégiques. Prenons l'exemple d'un contrôleur ONTAP avec un maximum de 500 000 IOPS et un mélange de charges de travail de production et de développement. Vous devez appliquer des règles de QoS maximales aux workloads de développement pour empêcher toute base de données de monopoliser le contrôleur. Vous appliqueriez ensuite des règles de QoS minimales aux charges de travail de production afin de vous assurer que les IOPS requises sont toujours disponibles, le cas échéant.

La QoS adaptative

La QoS adaptative fait référence à la fonctionnalité ONTAP où la limite de QoS repose sur la capacité de l'objet de stockage. Elle est rarement utilisée avec les bases de données, car il n'existe généralement aucun lien entre la taille d'une base de données et ses exigences de performances. Les grandes bases de données peuvent être quasiment inertes, tandis que les bases de données plus petites peuvent être celles qui nécessitent le plus d'IOPS.

La QoS adaptative peut s'avérer très utile avec les datastores de virtualisation, car les exigences en IOPS de ces jeux de données ont tendance à être corrélées à la taille totale de la base de données. Un datastore plus récent contenant 1 To de fichiers VMDK devrait avoir besoin d'environ la moitié des performances pour un datastore de 2 To. La QoS adaptative vous permet d'augmenter automatiquement les limites de qualité de service lorsque le datastore est rempli de données.

Efficacité

Les fonctionnalités ONTAP d'optimisation de l'espace sont optimisées pour les bases de données Oracle. Dans la plupart des cas, la meilleure approche consiste à conserver les valeurs par défaut avec toutes les fonctionnalités d'efficacité activées.

Les fonctionnalités d'optimisation de l'espace, telles que la compression, la compaction et la déduplication, sont conçues pour augmenter la quantité de données logiques correspondant à un volume de stockage physique donné. Vous réduisez ainsi vos coûts et vos frais de gestion.

À un niveau élevé, la compression est un processus mathématique qui permet de détecter et d'encoder des modèles de données de manière à réduire les besoins en espace. En revanche, la déduplication détecte les blocs de données répétés et supprime les copies parasites. La compaction permet à plusieurs blocs logiques de données de partager le même bloc physique sur le support.

> Reportez-vous aux sections ci-dessous sur le provisionnement fin pour une explication de l'interaction entre l'efficacité du stockage et la réservation fractionnaire.

Compression

Avant la disponibilité des systèmes de stockage 100 % Flash, la compression basée sur les baies était d'une valeur limitée, car la plupart des charges de travail exigeantes en E/S nécessitaient un très grand nombre de piles pour obtenir une performance acceptable. Les systèmes de stockage contenaient invariablement beaucoup plus de capacité que nécessaire, ce qui a pour effet d'augmenter le nombre de disques. La situation a changé avec la montée du stockage Solid-State. Il n'est plus nécessaire de surprovisionner des disques uniquement pour obtenir de bonnes performances. L'espace disque d'un système de stockage peut être adapté aux besoins réels en termes de capacité.

La capacité accrue des disques SSD en termes d'IOPS permet presque toujours de réaliser des économies

par rapport aux disques rotatifs. Toutefois, la compression peut réaliser davantage d'économies en augmentant la capacité effective des supports SSD.

Il existe plusieurs façons de compresser les données. De nombreuses bases de données incluent leurs propres fonctionnalités de compression, mais ce phénomène est rarement observé dans les environnements clients. La raison en est généralement la réduction des performances pour un **changement** de données compressées, plus avec certaines applications, il existe des coûts de licence élevés pour la compression au niveau de la base de données. Enfin, il y a les conséquences globales sur les performances des opérations des bases de données. Il est peu judicieux de payer un coût de licence par processeur élevé pour un processeur qui effectue la compression et la décompression des données plutôt que le véritable travail de base de données. Une meilleure option consiste à décharger la tâche de compression sur le système de stockage.

Compression adaptative

La compression adaptative a été testée en profondeur avec des charges de travail exigeantes sans effet sur les performances, même dans un environnement 100 % Flash où la latence se mesure en microsecondes. Certains clients ont même signalé une augmentation des performances due à l'utilisation de la compression, car les données restent compressées dans le cache, augmentant ainsi la quantité de cache disponible dans un contrôleur.

ONTAP gère les blocs physiques dans des unités de 4 Ko. La compression adaptative utilise une taille de bloc de compression par défaut de 8 Ko, ce qui signifie que les données sont compressées dans des unités de 8 Ko. La taille de bloc de 8 Ko la plus utilisée par les bases de données relationnelles est donc identique. Les algorithmes de compression deviennent plus efficaces avec la compression d'un volume croissant de données. Une taille de bloc de compression de 32 Ko serait plus compacte qu'une unité de bloc de compression de 8 Ko. Cela signifie que la compression adaptative utilisant une taille de bloc de 8 Ko par défaut entraîne des taux d'efficacité légèrement inférieurs, mais qu'une taille de bloc de compression inférieure présente également des avantages considérables. Les charges de travail de la base de données incluent une grande quantité d'activités de remplacement. Le remplacement d'un bloc de données de 32 Ko compressé de 8 Ko nécessite la lecture de l'intégralité des 32 Ko de données logiques, leur décompression, la mise à jour de la région de 8 Ko requise, la recompression, puis l'écriture de la totalité des 32 Ko sur les disques. Cette opération est très coûteuse pour un système de stockage. En effet, certaines baies de stockage concurrentes, basées sur des blocs de compression plus volumineux, affectent également considérablement les performances des charges de travail de la base de données.

> La taille de bloc utilisée par la compression adaptative peut être augmentée jusqu'à 32 Ko. Cela peut améliorer l'efficacité du stockage et doit être envisagé pour les fichiers de repos tels que les journaux de transactions et les fichiers de sauvegarde lorsqu'une quantité importante de ces données est stockée sur la baie. Dans certains cas, les bases de données actives qui utilisent une taille de bloc de 16 ou 32 Ko peuvent également tirer parti de l'augmentation de la taille de bloc de la compression adaptative pour qu'elle corresponde. Consultez un représentant NetApp ou partenaire pour savoir si cette solution convient à votre charge de travail.

> Les tailles de bloc de compression supérieures à 8 Ko ne doivent pas être utilisées avec la déduplication sur les destinations de sauvegarde en streaming. Les petites modifications apportées aux données sauvegardées affectent la fenêtre de compression de 32 Ko. Si la fenêtre change, les données compressées obtenues diffèrent dans l'ensemble du fichier. La déduplication a lieu après la compression, ce qui signifie que le moteur de déduplication voit chaque sauvegarde compressée différemment. Si la déduplication des sauvegardes en continu est nécessaire, seule une compression adaptative de bloc de 8 Ko doit être utilisée. Il est préférable d'utiliser la compression adaptative, car elle fonctionne à des blocs de taille réduite sans perturber l'efficacité de la déduplication. Pour des raisons similaires, la compression côté hôte interfère également avec l'efficacité de la déduplication.

Alignement de compression

La compression adaptative dans un environnement de base de données nécessite un certain respect de l'alignement des blocs de compression. Cela ne préoccupe que les données soumises à des écrasements aléatoires de blocs très spécifiques. Cette approche est similaire à l'alignement global du système de fichiers, où le début d'un système de fichiers doit être aligné sur une limite de périphérique de 4 Ko et la taille de bloc d'un système de fichiers doit être un multiple de 4 Ko.

Par exemple, une écriture de 8 Ko dans un fichier est compressée uniquement si elle s'aligne sur une limite de 8 Ko dans le système de fichiers lui-même. Ce point signifie qu'il doit figurer sur le premier 8 Ko du fichier, le deuxième 8 Ko du fichier, etc. La manière la plus simple de garantir un alignement correct est d'utiliser le type de LUN correct, toute partition créée doit avoir un décalage par rapport au début du périphérique qui est un multiple de 8K, et utiliser une taille de bloc du système de fichiers qui est un multiple de la taille de bloc de la base de données.

Les données telles que les sauvegardes ou les journaux de transactions sont des opérations écrites de manière séquentielle sur plusieurs blocs, qui sont tous compressés. Par conséquent, il n'est pas nécessaire de considérer l'alignement. Le seul modèle d'E/S préoccupant est l'écrasement aléatoire des fichiers.

Compaction

La compaction est une technologie qui améliore l'efficacité de la compression. Comme indiqué précédemment, la compression adaptative à elle seule permet d'économiser 2:1 au maximum, car elle se limite au stockage d'une E/S de 8 Ko dans un bloc WAFL de 4 Ko. Les méthodes de compression avec des blocs de taille supérieure améliorent l'efficacité. Cependant, elles ne conviennent pas aux données soumises à des remplacements de blocs de petite taille. La décompression d'unités de données de 32 Ko, la mise à jour d'une partie de 8 Ko, la recompression et l'écriture sur les disques entraînent une surcharge.

La compaction des données permet de stocker plusieurs blocs logiques dans des blocs physiques. Par exemple, une base de données avec des données fortement compressibles comme des blocs texte ou partiellement pleins peut être compressée de 8 Ko à 1 Ko. Sans compaction, 1 Ko de données occuperaient toujours un bloc complet de 4 Ko. La compaction des données à la volée permet de stocker 1 Ko de données compressées dans un espace physique de seulement 1 Ko, parallèlement à d'autres données compressées. Il ne s'agit pas d'une technologie de compression. Il s'agit simplement d'un moyen plus efficace d'allouer de l'espace sur les disques et, par conséquent, il ne doit pas créer d'effet détectable sur les performances.

Le degré d'économie obtenu varie. En général, les données déjà compressées ou chiffrées ne peuvent pas être compressées davantage et, par conséquent, la compaction de ces datasets ne peut pas être bénéfique. À contrario, les fichiers de données récemment initialisés ne contiennent qu'un petit peu plus que des métadonnées de bloc et des zéros compressent jusqu'à 80:1.

Efficacité du stockage sensible à la température

L'efficacité de stockage sensible à la température (TSSE) est disponible dans ONTAP 9.8 et versions ultérieures. Il s'appuie sur des cartes thermiques d'accès aux blocs pour identifier les blocs peu utilisés et les compresser à l'aide d'une meilleure efficacité.

Déduplication

La déduplication permet de supprimer les tailles de bloc dupliquées d'un dataset. Par exemple, si le même bloc de 4 Ko existe dans 10 fichiers différents, la déduplication redirige ce bloc de 4 Ko au sein des 10 fichiers vers le même bloc physique de 4 Ko. Résultat : une amélioration de l'efficacité de ces données de 10:1.

Les données, telles que les LUN de démarrage invité VMware, se dédupliquent extrêmement bien, car elles sont constituées de plusieurs copies des mêmes fichiers du système d'exploitation. L'efficacité de 100:1 et plus

ont été observées.

Certaines données ne contiennent pas de données dupliquées. Par exemple, un bloc Oracle contient un en-tête globalement unique à la base de données et une bande-annonce presque unique. Par conséquent, la déduplication d'une base de données Oracle permet rarement de réaliser plus de 1 % d'économies. La déduplication avec les bases de données MS SQL est légèrement meilleure, mais les métadonnées uniques au niveau des blocs restent une limitation.

Dans quelques cas, des économies d'espace allant jusqu'à 15 % ont été observées pour les bases de données de 16 Ko et les blocs volumineux. La bande de 4 Ko initiale de chaque bloc contient l'en-tête unique dans le monde, et le bloc de 4 Ko final contient la remorque presque unique. Les blocs internes sont candidats à la déduplication, bien que dans la pratique cela soit presque entièrement attribué à la déduplication des données mises à zéro.

De nombreuses baies concurrentes prétendent être capables de dédupliquer des bases de données en présumant qu'une base de données est copiée plusieurs fois. Il est également possible d'utiliser la déduplication NetApp, mais ONTAP offre une meilleure option : la technologie FlexClone de NetApp. Le résultat final est le même : plusieurs copies d'une base de données qui partagent la plupart des blocs physiques sous-jacents sont créées. L'utilisation de FlexClone est bien plus efficace que de prendre le temps de copier les fichiers de base de données, puis de les dédupliquer. Il s'agit en effet de la non-duplication plutôt que de la déduplication, car un doublon n'est jamais créé à la première place.

Efficacité et provisionnement fin

Les fonctions d'efficacité sont des formes de provisionnement fin. Par exemple, une LUN de 100 Go occupant un volume de 100 Go peut compresser à 50 Go. Aucune économie réelle n'est encore réalisée, car le volume est toujours de 100 Go. Le volume doit d'abord être réduit afin que l'espace économisé puisse être utilisé ailleurs sur le système. Si des modifications ultérieures de la LUN de 100 Go réduisent la taille des données compressibles, la LUN augmente et le volume pourrait se remplir.

Le provisionnement fin est fortement recommandé car il simplifie la gestion tout en améliorant la capacité exploitable avec les économies associées. La raison en est simple : les environnements de base de données comportent souvent beaucoup d'espace vide, un grand nombre de volumes et de LUN, ainsi que des données compressibles. Le provisionnement fin entraîne la réservation d'espace sur le stockage pour les volumes et les LUN au cas où un jour ils se traduirait par une saturation de 100 % et contiendraient des données non compressibles à 100 %. Il est peu probable que cela se produise. Le provisionnement fin permet de récupérer et d'utiliser cet espace ailleurs. Il permet également de gérer la capacité en fonction du système de stockage lui-même, plutôt que de nombreux volumes et LUN plus petits.

Certains clients préfèrent utiliser le provisionnement lourd, soit pour des charges de travail spécifiques, soit généralement en fonction de pratiques opérationnelles et d'approvisionnement établies.

> Si un volume est configuré en mode « Thick provisioning », veillez à désactiver complètement toutes les fonctions d'efficacité de ce volume, y compris la décompression et la suppression de la déduplication à l'aide de la `sis undo` commande. Le volume ne doit pas apparaître en `volume efficiency show` sortie. Si c'est le cas, le volume est encore partiellement configuré pour les fonctions d'efficacité. Par conséquent, les garanties de remplacement fonctionnent différemment, ce qui augmente le risque que les dépassements de configuration entraînent un manque inattendu d'espace du volume, ce qui entraîne des erreurs d'E/S de la base de données.

Meilleures pratiques en matière d'efficacité

NetApp recommande ce qui suit :

AFF par défaut

Les volumes créés sur ONTAP et exécutés sur un système AFF 100 % Flash sont à allocation dynamique, avec l'activation de toutes les fonctionnalités d'efficacité à la volée. Bien que les bases de données ne bénéficient généralement pas de la déduplication et puissent inclure des données non compressibles, les paramètres par défaut conviennent néanmoins à la plupart des charges de travail. ONTAP est conçu pour traiter efficacement tous les types de données et de modèles d'E/S, qu'ils entraînent ou non des économies. Les valeurs par défaut ne doivent être modifiées que si les raisons sont parfaitement comprises et si un écart est bénéfique.

Recommandations générales

- Si les volumes et/ou les LUN ne sont pas à provisionnement fin, vous devez désactiver tous les paramètres d'efficacité car l'utilisation de ces fonctionnalités ne permet aucune économie et la combinaison du provisionnement lourd et de l'optimisation de l'espace peut provoquer des comportements inattendus, notamment des erreurs de manque d'espace.
- Si les données ne sont pas sujettes à des écrasements, par exemple avec des sauvegardes ou des journaux de transactions de base de données, vous pouvez atteindre une meilleure efficacité en activant TSSE avec une période de refroidissement faible.
- Certains fichiers peuvent contenir une quantité importante de données non compressibles, par exemple lorsque la compression est déjà activée au niveau de l'application, les fichiers sont cryptés. Si l'un de ces scénarios est vrai, envisagez de désactiver la compression pour permettre un fonctionnement plus efficace sur d'autres volumes contenant des données compressibles.
- N'utilisez pas la compression et la déduplication de 32 Ko pour les sauvegardes de bases de données. Voir la section Compression adaptative pour plus d'informations.

Provisionnement fin

Le provisionnement fin pour une base de données Oracle nécessite une planification minutieuse, car il en résulte une configuration d'espace sur un système de stockage qui n'est pas nécessairement physiquement disponible. Cela vaut vraiment le coup, car une fois correctement effectué, il en résulte des économies considérables et des améliorations en termes de gestion.

Le provisionnement fin, de nombreuses formes, fait partie intégrante de nombreuses fonctionnalités offertes par ONTAP à l'environnement applicatif d'entreprise. Le provisionnement fin est également étroitement lié aux technologies d'efficacité pour la même raison : les fonctionnalités d'efficacité permettent de stocker davantage de données logiques que ce qui existe techniquement sur le système de stockage.

La plupart des snapshots impliquent un provisionnement fin. Par exemple, une base de données classique de 10 To sur un système de stockage NetApp compte environ 30 jours de copies Snapshot. Cet arrangement donne lieu à environ 10 To de données visibles dans le système de fichiers actif et 300 To dédiés aux snapshots. La capacité totale de stockage de 310 To réside généralement dans un espace d'environ 12 To à 15 To. La base de données active consomme 10 To et les 300 To de données restantes ne nécessitent que 2 à 5 To d'espace, car seules les modifications apportées aux données d'origine sont stockées.

Le clonage est également un exemple de provisionnement fin. Un client NetApp majeur a créé 40 clones d'une base de données de 80 To à utiliser pour le développement. Si les 40 développeurs qui utilisent ces clones surécrivent chaque bloc dans chaque fichier de données, plus de 3,2 po de stockage seraient nécessaires. En pratique, le chiffre d'affaires est faible et l'espace collectif requis est proche de 40 To, car seules les modifications sont stockées sur les disques.

Gestion de l'espace

Le provisionnement fin d'un environnement applicatif doit être extrêmement prudent, car les taux de modification des données peuvent augmenter de manière inattendue. Par exemple, la consommation d'espace due aux snapshots peut augmenter rapidement si les tables de base de données sont réindexées ou si des correctifs à grande échelle sont appliqués aux invités VMware. Une sauvegarde mal placée peut écrire une grande quantité de données dans un délai très court. Enfin, il peut être difficile de restaurer certaines applications si un système de fichiers manque d'espace de façon inattendue.

Avec une configuration soigneuse de, ces risques peuvent être maîtrisés `volume-autogrow` et `snapshot-autodelete` règles. Comme leurs noms l'indiquent, ces options permettent de créer des règles qui effacent automatiquement l'espace consommé par les snapshots ou augmentent un volume pour prendre en charge des données supplémentaires. De nombreuses options sont disponibles et les besoins varient selon les clients.

Voir la "documentation sur la gestion du stockage logique" pour une discussion complète de ces fonctionnalités.

Réservations fractionnaires

La réserve fractionnaire fait référence au comportement d'une LUN dans un volume en ce qui concerne l'efficacité de l'espace. Lorsque l'option `fractional-reserve` est défini sur 100 %. toutes les données du volume peuvent connaître un taux de rotation de 100 % avec n'importe quel modèle de données, sans épuiser l'espace sur le volume.

Par exemple, prenons l'exemple d'une base de données située sur une seule LUN de 250 Go dans un volume de 1 To. La création d'un snapshot entraînerait immédiatement la réservation d'un espace supplémentaire de 250 Go dans le volume, garantissant ainsi que l'espace disponible sur le volume ne serait pas insuffisant pour quelque raison que ce soit. L'utilisation de réserves fractionnaires est généralement inutile car il est très peu probable que chaque octet du volume de base de données ait besoin d'être écrasé. Il n'y a aucune raison de réserver de l'espace pour un événement qui ne se produit jamais. Cependant, si un client ne peut pas surveiller la consommation d'espace dans un système de stockage et doit être certain que l'espace ne sera jamais épuisé, des réservations fractionnaires de 100 % seront nécessaires pour utiliser les snapshots.

Compression et déduplication

La compression et la déduplication sont deux formes de provisionnement fin. Par exemple, une empreinte des données de 50 To peut être compressée jusqu'à 30 To, ce qui permet d'économiser 20 To. Pour que la compression offre tous les avantages, il faut utiliser quelques 20 To pour d'autres données ou acheter le système de stockage avec moins de 50 To. Il en résulte une quantité de données stockées supérieure à ce qui n'est techniquement disponible sur le système de stockage. Du point de vue des données, il y a 50 To de données, même si celles-ci ne occupent que 30 To sur les disques.

Il est toujours possible que la compressibilité d'un dataset change, ce qui entraîne une consommation accrue de l'espace réel. Cette augmentation de la consommation signifie que la compression doit être gérée comme avec les autres formes de provisionnement fin en termes de surveillance et d'utilisation `volume-autogrow` et `snapshot-autodelete`.

La compression et la déduplication sont présentées plus en détail dans la section xref:./oracle/efficiency.html

Compression et réservations fractionnaires

La compression est une forme d'allocation dynamique. Les réservations fractionnaires affectent l'utilisation de la compression, avec une remarque importante ; l'espace est réservé avant la création du snapshot. Normalement, la réserve fractionnaire n'est importante que si un instantané existe. S'il n'y a pas de snapshot,

la réserve fractionnaire n'est pas importante. Ce n'est pas le cas avec la compression. Si une LUN est créée sur un volume avec compression, ONTAP conserve l'espace nécessaire pour prendre en charge un snapshot. Ce comportement peut être déroutant pendant la configuration, mais il est normal.

Prenons l'exemple d'un volume de 10 Go avec une LUN de 5 Go compressée à 2,5 Go sans copie Snapshot. Prenez en compte ces deux scénarios :

- La réserve fractionnaire = 100 entraîne une utilisation de 7,5 Go
- La réserve fractionnaire = 0 entraîne une utilisation de 2,5 Go

Le premier scénario comprend 2,5 Go de consommation d'espace pour les données actuelles et 5 Go d'espace pour représenter 100 % de chiffre d'affaires de la source en prévision de l'utilisation des snapshots. Le deuxième scénario ne réserve pas d'espace supplémentaire.

Bien que cette situation puisse sembler confuse, il est peu probable qu'elle soit rencontrée dans la pratique. La compression implique un provisionnement fin et le provisionnement fin dans un environnement LUN nécessite des réservations fractionnaires. Il est toujours possible d'écraser des données compressées par un élément non compressible, ce qui signifie qu'un volume doit être à provisionnement fin pour la compression, pour réaliser des économies.

NetApp recommande les configurations de réserve suivantes :

- Réglez `fractional-reserve` à 0 lorsque la surveillance de la capacité de base est en place avec `volume-autogrow` et `snapshot-autodelete`.
- Réglez `fractional-reserve` à 100 s'il n'y a pas de capacité de surveillance ou s'il est impossible d'évacuer l'espace en quelque circonstance que ce soit.

Allocation d'espace libre et d'espace LVM

L'efficacité du provisionnement fin des LUN actives dans un environnement de système de fichiers peut être perdue au fil du temps suite à la suppression des données. À moins que les données supprimées ne soient écrasées par des zéros (voir également "ASMRU" Ou l'espace est libéré avec la récupération d'espace TRIM/UNMAP, les données « effacées » occupent de plus en plus d'espace non alloué dans le système de fichiers. En outre, l'utilisation du provisionnement fin des LUN actives est limitée dans de nombreux environnements de base de données, car les fichiers de données sont initialisés sur leur taille complète au moment de la création.

Une planification minutieuse de la configuration de LVM peut améliorer l'efficacité et réduire les besoins en provisionnement du stockage et en redimensionnement des LUN. Lorsqu'un LVM tel que Veritas VxVM ou Oracle ASM est utilisé, les LUN sous-jacentes sont divisés en extensions qui ne sont utilisées que lorsque cela est nécessaire. Par exemple, si un dataset commence à 2 To mais peut atteindre 10 To au fil du temps, ce dataset peut être placé sur 10 To de LUN à provisionnement fin organisées dans un groupe de disques LVM. Elle occupant seulement 2 To d'espace au moment de la création et réclarait uniquement de l'espace supplémentaire, dans la mesure où les extensions sont allouées pour prendre en charge la croissance du volume des données. Ce processus est sûr tant que l'espace est surveillé.

Basculement/basculement ONTAP

Il est nécessaire de bien comprendre les fonctions de basculement et de basculement du stockage pour s'assurer que les opérations de la base de données Oracle ne sont pas interrompues par ces opérations. En outre, les arguments utilisés par les opérations de basculement et de basculement peuvent affecter l'intégrité des données en cas

d’utilisation incorrecte.

- Dans des conditions normales, les écritures entrantes sur un contrôleur donné sont mises en miroir de manière synchrone sur son partenaire. Dans un environnement NetApp MetroCluster, les écritures sont également mises en miroir sur un contrôleur distant. Tant qu’une écriture n’est pas stockée sur un support non volatile dans tous les emplacements, elle n’est pas validée par l’application hôte.
- Le support qui stocke les données d’écriture est appelé mémoire non volatile ou NVMEM. Elle est également parfois appelée mémoire NVRAM, et peut être considérée comme un cache d’écriture, même si elle fonctionne comme un journal. En fonctionnement normal, les données de NVMEM ne sont pas lues ; elles sont uniquement utilisées pour protéger les données en cas de défaillance logicielle ou matérielle. Lors de l’écriture des données sur les disques, les données sont transférées de la mémoire RAM du système, et non de NVMEM.
- Lors d’une opération de basculement, un nœud d’une paire haute disponibilité reprend les opérations de son partenaire. Un basculement est quasiment identique, mais s’applique aux configurations MetroCluster dans lesquelles un nœud distant prend le relais par rapport à un nœud local.

Lors des opérations de maintenance de routine, un basculement du stockage ou un basculement doivent être transparents, sauf en cas de brève pause potentielle dans les opérations en cas de changement des chemins réseau. La mise en réseau peut toutefois être complexe et il est facile d’y faire des erreurs. NetApp recommande donc de tester minutieusement les opérations de basculement et de basculement avant de mettre en production un système de stockage. C’est la seule façon de s’assurer que tous les chemins réseau sont correctement configurés. Dans un environnement SAN, vérifiez soigneusement le résultat de la commande `sanlun lun show -p` pour vous assurer que tous les chemins principaux et secondaires attendus sont disponibles.

Il convient de faire attention lors d’un basculement forcé ou d’un basculement forcé. Forcer une modification de la configuration du stockage avec ces options signifie que l’état du contrôleur propriétaire des disques est ignoré et que le nœud alternatif prend le contrôle des disques. Une force de basculement incorrecte peut entraîner une perte ou une corruption des données. En effet, un basculement forcé ou un basculement forcé peut rejeter le contenu de la NVMEM. Une fois le basculement ou le basculement effectué, la perte de ces données signifie que les données stockées sur les disques peuvent revenir à un état plus ancien du point de vue de la base de données.

Un basculement forcé avec une paire haute disponibilité normale devrait rarement être nécessaire. Dans la plupart des scénarios de défaillance, un nœud s’arrête et informe le partenaire qu’un basculement automatique a lieu. Il existe certains cas à la périphérie, par exemple une panne de déploiement où l’interconnexion entre les nœuds est perdue puis un contrôleur est perdu, dans lequel un basculement forcé est nécessaire. Dans ce cas, la mise en miroir entre les nœuds est perdue avant la panne du contrôleur, ce qui signifie que le contrôleur survivant n’aurait plus de copie des écritures en cours. Le basculement doit ensuite être forcé, ce qui signifie que des données peuvent être perdues.

La même logique s’applique à un basculement MetroCluster. Dans des conditions normales, le basculement est presque transparent. Toutefois, un incident peut entraîner une perte de connectivité entre le site survivant et le site de reprise sur incident. Du point de vue du site survivant, le problème ne pourrait être rien de plus qu’une interruption de la connectivité entre les sites, et le site d’origine pourrait encore traiter les données. Si un nœud ne peut pas vérifier l’état du contrôleur principal, seul un basculement forcé est possible.

NetApp recommande de prendre les précautions suivantes :

- Veillez à ne pas forcer accidentellement un basculement ou un basculement. En règle générale, il n'est pas nécessaire de forcer et le fait de forcer la modification peut entraîner la perte de données.
- Si un basculement ou un basculement forcé s'avère nécessaire, assurez-vous que les applications sont arrêtées, que tous les systèmes de fichiers sont démontés et que les groupes de volumes LVM (Logical Volume Manager) sont proposés en mode Variyoffed. Les groupes de disques ASM doivent être démontés.
- En cas de basculement forcé du MetroCluster, vous pouvez isoler le nœud défaillant de toutes les ressources de stockage restantes. Pour plus d'informations, consultez le Guide de gestion et de reprise sur incident de MetroCluster correspondant à la version appropriée de ONTAP.

MetroCluster et plusieurs agrégats

MetroCluster est une technologie de réplication synchrone qui passe en mode asynchrone en cas d'interruption de la connectivité. Cette demande est la plus courante de la part des clients, car une réplication synchrone garantie signifie que l'interruption de la connectivité du site entraîne un blocage complet des E/S de la base de données, ce qui la met hors service.

Avec MetroCluster, les agrégats sont rapidement resynchronisés une fois la connectivité restaurée. Contrairement à d'autres technologies de stockage, MetroCluster ne devrait jamais nécessiter de mise en miroir complète après une panne de site. Seules les modifications delta doivent être expédiées.

Dans les jeux de données qui couvrent les agrégats, le risque est faible de nécessiter des étapes supplémentaires de restauration des données en cas de sinistre en cas de déploiement. En particulier, si (a) la connectivité entre les sites est interrompue, (b) la connectivité est restaurée, (c) les agrégats atteignent un état dans lequel certains sont synchronisés et d'autres ne le sont pas, puis (d) le site primaire est perdu, le site survivant dans lequel les agrégats ne sont pas synchronisés. Dans ce cas, une partie du dataset est synchronisée et il est impossible d'ouvrir des applications, des bases de données ou des datastores sans restauration. Si un dataset compte plusieurs agrégats, NetApp recommande vivement d'utiliser des sauvegardes basées sur des snapshots avec l'un des nombreux outils disponibles pour vérifier la restauration rapide dans ce scénario inhabituel.

Configuration de la base de données

Tailles de bloc

ONTAP utilise en interne une taille de bloc variable, ce qui signifie que les bases de données Oracle peuvent être configurées avec n'importe quelle taille de bloc. Cependant, la taille des blocs du système de fichiers peut affecter les performances et, dans certains cas, une taille de bloc de reprise supérieure peut améliorer les performances.

Tailles des blocs de fichiers de données

Certains systèmes d'exploitation offrent un choix de tailles de blocs de système de fichiers. Pour les systèmes de fichiers prenant en charge les fichiers de données Oracle, la taille de bloc doit être de 8 Ko lorsque la compression est utilisée. Lorsque la compression n'est pas requise, vous pouvez utiliser une taille de bloc de 8 Ko ou 4 Ko.

Si un fichier de données est placé sur un système de fichiers avec un bloc de 512 octets, des fichiers mal alignés sont possibles. Il est possible que le LUN et le système de fichiers soient correctement alignés en fonction des recommandations de NetApp, mais les E/S de fichier sont mal alignées. Un tel mauvais alignement entraînerait de graves problèmes de performances.

Les systèmes de fichiers prenant en charge les journaux de reprise doivent utiliser une taille de bloc qui représente un multiple de la taille de bloc de reprise. Cela nécessite généralement que le système de fichiers redo log et le fichier redo log lui-même utilisent une taille de bloc de 512 octets.

Rétablir les tailles des blocs

Avec des taux de reprise très élevés, il est possible que des tailles de bloc de 4 Ko soient plus performantes, car les taux de reprise élevés permettent d'exécuter les E/S en moins d'opérations et de manière plus efficace. Si les taux de reprise sont supérieurs à 50 Mbit/s, envisagez de tester une taille de bloc de 4 Ko.

Quelques problèmes clients ont été identifiés avec les bases de données à l'aide de journaux de reprise avec une taille de bloc de 512 octets sur un système de fichiers d'une taille de bloc de 4 Ko et de nombreuses transactions très petites. La surcharge liée à l'application de plusieurs modifications de 512 octets à un seul bloc du système de fichiers de 4 Ko a entraîné des problèmes de performances qui ont été résolus en changeant le système de fichiers pour qu'il utilise une taille de bloc de 512 octets.

NetApp vous recommande de ne pas modifier la taille du bloc de reprise, sauf si un service client ou un service professionnel vous en informe ou si le changement est basé sur la documentation officielle du produit.

db_file_multibloc_read_count

Le `db_file_multiblock_read_count` Paramètre contrôle le nombre maximal de blocs de base de données Oracle lus par Oracle au cours d'une opération, pendant les E/S séquentielles

Toutefois, ce paramètre n'affecte pas le nombre de blocs lus par Oracle au cours des opérations de lecture, ni le nombre d'E/S aléatoires Seule la taille de bloc des E/S séquentielles est affectée.

Oracle recommande à l'utilisateur de ne pas définir ce paramètre. Cela permet au logiciel de base de données de définir automatiquement la valeur optimale. Cela signifie généralement que ce paramètre est défini sur une valeur qui produit une taille d'E/S de 1 Mo. Par exemple, une lecture de 1 Mo de blocs de 8 Ko nécessite la lecture de 128 blocs. La valeur par défaut de ce paramètre est donc 128.

La plupart des problèmes de performance de base de données observés par NetApp sur les sites des clients provenaient de paramètres incorrects. Des raisons valides ont été données pour modifier cette valeur avec les versions 8 et 9 d'Oracle. Par conséquent, le paramètre peut être présent sans le savoir dans `init.ora` Fichiers car la base de données a été mise à niveau vers Oracle 10 et versions ultérieures. La configuration héritée de 8 ou 16, par rapport à la valeur par défaut 128, nuit de manière significative aux performances d'E/S séquentielles.

NetApp recommande de régler le `db_file_multiblock_read_count` le paramètre ne doit pas être présent dans le `init.ora` fichier. NetApp n'a jamais observé d'amélioration des performances suite à la modification de ce paramètre, mais le débit d'E/S séquentielles subit une importante dégradation dans de nombreux cas.

filesytemio_options

Le paramètre d'initialisation Oracle `filesystemio_options` Contrôle l'utilisation des E/S asynchrones et directes

Contrairement à une idée reçue, ces deux types d'E/S ne s'excluent pas mutuellement. NetApp a observé que ce paramètre est souvent mal configuré dans les environnements des clients. Cette configuration incorrecte est la cause directe de nombreux problèmes de performances.

Les E/S asynchrones offrent la possibilité de paralléliser les opérations Oracle d'E/S. Avant la disponibilité des E/S asynchrones sur différents systèmes d'exploitation, les utilisateurs ont configuré de nombreux processus dbwriter et modifié la configuration du processus serveur. Avec les E/S asynchrones, le système d'exploitation lui-même exécute les E/S en parallèle pour le compte du logiciel de base de données. Ce processus ne présente aucun risque pour les données et les opérations critiques, telles que la journalisation de reprise Oracle, sont toujours exécutées de manière synchrone.

Les E/S directes contournent le cache du tampon du système d'exploitation. Sur un système UNIX, les E/S transitent normalement par le cache du tampon du système d'exploitation. Ceci est utile pour les applications qui ne maintiennent pas de cache interne, mais Oracle dispose de son propre cache de tampon dans la SGA. Dans la plupart des cas, il est préférable d'activer les E/S directes et d'allouer la RAM du serveur à la mémoire SGA plutôt que d'utiliser le cache du tampon du système d'exploitation. La SGA exploite la mémoire plus efficacement. En outre, lors de leur transit via le tampon du se, les E/S sont soumises à un traitement supplémentaire, ce qui augmente les latences. Cette augmentation est particulièrement visible lors des E/S intenses en écriture, pour lesquelles la faible latence est primordiale.

Les options pour `filesystemio_options` sont :

- **Async.** Oracle soumet des demandes d'E/S au système d'exploitation pour traitement. Ce qui lui permet d'effectuer d'autres tâches plutôt que d'attendre la fin des E/S et d'augmenter ainsi la parallélisation des E/S.
- **Directio.** Oracle effectue des E/S directement par rapport aux fichiers physiques plutôt que de router les E/S via le cache du système d'exploitation hôte.
- **None.** Oracle utilise des E/S synchrones et mises en tampon Dans cette configuration, le choix entre les processus serveur partagés et dédiés et le nombre de dbwriter est plus important.
- **Setall.** Oracle utilise des E/S asynchrones et directes Dans presque tous les cas, l'utilisation de `setall` est optimale.

> Le `filesystemio_options` Ce paramètre n'a aucun effet dans les environnements dNFS et ASM. Dans ces environnements, les E/S asynchrones et directes sont automatiquement utilisées

Certains clients ont déjà rencontré des problèmes d'E/S asynchrones, notamment avec les versions précédentes de Red Hat Enterprise Linux 4 (RHEL4). Certains conseils obsolètes sur Internet suggèrent toujours d'éviter les E/S asynchrones en raison d'informations obsolètes. Les E/S asynchrones sont stables sur tous les systèmes d'exploitation actuels. Il n'y a aucune raison de le désactiver, en l'absence d'un bug connu avec le système d'exploitation.

Si une base de données utilise des E/S mises en tampon, un switch vers des E/S directes peut également justifier une modification de la taille de la mémoire SGA. La désactivation des E/S mises en tampon élimine le gain de performance fourni par le cache du se hôte pour la base de données. L'ajout de RAM à la SGA résout ce problème. Et devrait améliorer les performances nettes d'E/S.

Bien qu'il soit presque toujours préférable d'utiliser la RAM pour la SGA d'Oracle plutôt que pour le cache du tampon du système d'exploitation, il peut s'avérer impossible de déterminer ce qui est le plus avantageux. Par exemple, il est parfois préférable d'utiliser des E/S mises en tampon avec une mémoire SGA de très petite taille sur un serveur de base de données comportant de nombreuses instances Oracle actives par intermittence. Cette configuration permet à toutes les instances de base de données en cours d'exécution d'utiliser de manière flexible la RAM restante sur le système d'exploitation. Cette situation est très inhabituelle, mais elle a été observée sur certains sites clients.

NetApp recommande le réglage `filesystemio_options` à `setall`, Mais notez que dans certains cas, la perte du cache du tampon hôte peut nécessiter une augmentation de la SGA d'Oracle.

Délais d'expiration du RAC

Oracle RAC est un produit clusterware qui comporte plusieurs types de processus de pulsation internes qui contrôlent l'intégrité du cluster.

Les informations dans le "misscount" La section contient des informations essentielles pour les environnements RAC Oracle utilisant un stockage en réseau. Dans la plupart des cas, les paramètres RAC Oracle par défaut devront être modifiés pour garantir que le cluster RAC résiste aux modifications de chemin réseau et aux opérations de basculement/basculement du stockage.

disktimeout

Le paramètre RAC principal lié au stockage est `disktimeout`. Ce paramètre contrôle le seuil au sein duquel les E/S du fichier de vote doivent être terminées. Si le `disktimeout` Le paramètre est dépassé, puis le nœud RAC est supprimé du cluster. La valeur par défaut de ce paramètre est 200. Cette valeur doit être suffisante pour les procédures standard de Takeover et and Giveback du stockage.

NetApp recommande fortement de tester soigneusement les configurations RAC avant de les mettre en production, car de nombreux facteurs affectent un basculement ou un rétablissement. Outre le temps nécessaire au basculement du stockage, la propagation des modifications du protocole LACP (Link Aggregation Control Protocol) nécessite également du temps supplémentaire. En outre, le logiciel de chemins d'accès multiples SAN doit détecter un délai d'expiration d'E/S et réessayer sur un autre chemin. Si une base de données est extrêmement active, une grande quantité d'E/S doit être mise en file d'attente et relancée avant le traitement des E/S du disque de vote.

En l'absence d'un basculement ou d'un retour de stockage réel, l'effet peut être simulé à l'aide de tests de câble Pull sur le serveur de base de données.

NetApp recommande ce qui suit :

- En quittant le `disktimeout` paramètre à la valeur par défaut de 200.
- Testez toujours soigneusement une configuration RAC.

misscount

Le `misscount` Le paramètre affecte normalement uniquement la pulsation réseau entre les nœuds RAC. La valeur par défaut est 30 secondes. Si les binaires de la grille se trouvent sur une matrice de stockage ou si le disque d'amorçage du système d'exploitation n'est pas local, ce paramètre peut devenir important. Cela inclut

les hôtes avec des lecteurs de démarrage situés sur un SAN FC, les systèmes d'exploitation démarrés par NFS et les lecteurs de démarrage situés sur les datastores de virtualisation, tels qu'un fichier VMDK.

Si l'accès à un disque de démarrage est interrompu par un basculement ou un rétablissement du stockage, il est possible que l'emplacement binaire de la grille ou l'ensemble du système d'exploitation soit temporairement bloqué. Le temps nécessaire à ONTAP pour terminer l'opération de stockage et au système d'exploitation pour changer les chemins et reprendre les E/S peut être supérieur à `misscount` seuil. Par conséquent, un nœud est immédiatement supprimé une fois la connectivité à la LUN de démarrage ou aux binaires de la grille restaurée. Dans la plupart des cas, l'exclusion et le redémarrage qui s'ensuit se produisent sans message de journalisation indiquant la raison du redémarrage. Toutes les configurations ne sont pas affectées. Testez donc tout hôte de démarrage SAN, de démarrage NFS ou basé sur un datastore dans un environnement RAC afin que RAC reste stable si la communication avec le lecteur de démarrage est interrompue.

Dans le cas de lecteurs de démarrage non locaux ou d'un système de fichiers non local hébergeant `grid` binaires, le `misscount` devra être modifié pour correspondre `disktimeout`. Si ce paramètre est modifié, effectuez des tests supplémentaires pour identifier également les effets sur le comportement du RAC, tels que le temps de basculement du nœud.

NetApp recommande ce qui suit :

- Quittez le `misscount` paramètre à la valeur par défaut de 30, sauf si l'une des conditions suivantes s'applique :
  - `grid` Les fichiers binaires sont situés sur un disque connecté au réseau, y compris les disques basés sur NFS, iSCSI, FC et les datastores.
  - Le système d'exploitation est démarré sur un SAN.
- Dans de tels cas, évaluez l'effet des interruptions de réseau qui affectent l'accès au système d'exploitation ou `GRID_HOME` systèmes de fichiers. Dans certains cas, de telles interruptions provoquent le blocage des démons RAC Oracle, ce qui peut conduire à un `misscount`délai d'expiration et suppression basés sur. Le délai par défaut est de 27 secondes, soit la valeur de `misscount` moins `reboottime`. Dans de tels cas, augmenter `misscount` à 200 pour correspondre `disktimeout`.

Configuration de l'hôte

AIX

Rubriques de configuration pour la base de données Oracle sous IBM AIX avec ONTAP.

E/S simultanées

Pour obtenir des performances optimales sur IBM AIX, il est nécessaire d'utiliser des E/S simultanées Sans E/S simultanées, les limites de performances sont probablement dues au fait qu'AIX exécute des E/S atomiques sérialisées, ce qui entraîne une surcharge importante.

À l'origine, NetApp a recommandé d'utiliser le `cio` Option de montage pour forcer l'utilisation d'E/S simultanées sur le système de fichiers, mais ce processus présente des inconvénients et n'est plus nécessaire. Depuis l'introduction d'AIX 5.2 et d'Oracle 10gR1, Oracle sous AIX peut ouvrir des fichiers individuels pour des E/S simultanées, au lieu de forcer des E/S simultanées sur l'ensemble du système de fichiers.

La meilleure méthode pour activer les E/S simultanées est de définir le `init.ora` paramètre `filesystemio_options` à `setall`. Oracle peut ainsi ouvrir des fichiers spécifiques pour une utilisation avec des E/S simultanées

À l'aide de `cio` En tant qu'option de montage, force l'utilisation d'E/S simultanées, ce qui peut avoir des conséquences négatives. Par exemple, forcer des E/S simultanées désactive la lecture anticipée sur les systèmes de fichiers, ce qui peut nuire aux performances des E/S se produisant en dehors du logiciel de base de données Oracle, comme la copie de fichiers et les sauvegardes sur bande. En outre, les produits tels qu'Oracle GoldenGate et SAP BR*Tools ne sont pas compatibles avec l'utilisation du `cio` Option de montage avec certaines versions d'Oracle.

NetApp recommande ce qui suit :

- N'utilisez pas le `cio` option de montage au niveau du système de fichiers. Activez plutôt les E/S simultanées via l'utilisation de `filesystemio_options=setall`.
- Utilisez uniquement le `cio` l'option de montage doit être définie si elle n'est pas possible `filesystemio_options=setall`.

Options de montage NFS AIX

Le tableau suivant répertorie les options de montage NFS AIX pour les bases de données Oracle à instance unique.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, intr`

Le tableau suivant répertorie les options de montage NFS AIX pour RAC.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac`
`CRS/Voting`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac`
Ressource dédiée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`

Type de fichier	Options de montage
Partagée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr`

La principale différence entre les options de montage à instance unique et RAC est l'ajout de `noac` aux options de montage. Cet ajout a pour effet de désactiver la mise en cache du système d'exploitation hôte qui permet à toutes les instances du cluster RAC d'avoir une vue cohérente de l'état des données.

En utilisant le `cio` option de montage et `init.ora` paramètre `filesystemio_options=setall` a le même effet que la désactivation de la mise en cache de l'hôte, il est toujours nécessaire de l'utiliser `noac`. `noac` est requis pour le partage `ORACLE_HOME` Déploiements pour faciliter la cohérence des fichiers tels que les fichiers de mots de passe Oracle et `spfile` fichiers de paramètres. Si chaque instance d'un cluster RAC possède un dédié `ORACLE_HOME`, ce paramètre n'est pas requis.

Options de montage AIX jfs/jfs2

Le tableau suivant répertorie les options de montage AIX jfs/jfs2.

Type de fichier	Options de montage
Accueil ADR	Valeurs par défaut
Fichiers de contrôle Fichiers de données Journaux de reprise	Valeurs par défaut
ORACLE_HOME	Valeurs par défaut

Avant d'utiliser AIX `hdisk` dans tout environnement, y compris les bases de données, vérifiez le paramètre `queue_depth`. Ce paramètre n'est pas la profondeur de la file d'attente HBA ; il se rapporte plutôt à la profondeur de la file d'attente SCSI de l'individu `hdisk device. Depending on how the LUNs are configured, the value for `queue_depth` peut être trop faible pour de bonnes performances. Les tests ont montré que la valeur optimale est de 64.

HP-UX

Rubriques de configuration pour la base de données Oracle sur HP-UX avec ONTAP.

Options de montage NFS HP-UX

Le tableau suivant répertorie les options de montage NFS HP-UX pour une seule instance.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144,forc edirectio, nointr,suid`

Type de fichier	Options de montage
`ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`

Le tableau suivant répertorie les options de montage NFS HP-UX pour RAC.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, noac,suid`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw, bg,hard, [vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,forcedirectio,suid`
CRS/vote	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac, forcedirectio,suid`
Ressource dédiée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid`
Partagée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,suid`

La principale différence entre les options de montage à instance unique et RAC est l'ajout de `noac` et `forcedirectio` aux options de montage. Cet ajout a pour effet de désactiver la mise en cache du système d'exploitation hôte, ce qui permet à toutes les instances du cluster RAC d'avoir une vue cohérente de l'état des données. En utilisant le `init.ora` paramètre `filesystemio_options=setall` a le même effet que la désactivation de la mise en cache de l'hôte, il est toujours nécessaire de l'utiliser `noac` et `forcedirectio`.

La raison `noac` est requis pour le partage `ORACLE_HOME` Les déploiements visent à faciliter la cohérence des fichiers tels que les fichiers de mots de passe Oracle et les fichiers spfiles. Si chaque instance d'un cluster RAC possède un dédié `ORACLE_HOME`, ce paramètre n'est pas requis.

Options de montage HP-UX VxFS

Utilisez les options de montage suivantes pour les systèmes de fichiers hébergeant les binaires Oracle :

```
delaylog,nodatainlog
```

Utilisez les options de montage suivantes pour les systèmes de fichiers contenant des fichiers de données, des journaux de reprise, des journaux d'archivage et des fichiers de contrôle dans lesquels la version de HP-UX ne prend pas en charge les E/S simultanées :

```
nodatainlog,mincache=direct,convosync=direct
```

Lorsque des E/S simultanées sont prises en charge (VxFS 5.0.1 et versions ultérieures, ou avec ServiceGuard Storage Management Suite), utilisez ces options de montage pour les systèmes de fichiers contenant des fichiers de données, des journaux de reprise, des journaux d'archivage et des fichiers de contrôle :

```
delaylog,cio
```

> Le paramètre `db_file_multiblock_read_count` Est particulièrement critique dans les environnements VxFS. Oracle recommande que ce paramètre ne soit pas défini dans Oracle 10g R1 et versions ultérieures, sauf indication contraire. La taille de bloc Oracle de 8 Ko par défaut est 128. Si la valeur de ce paramètre est forcée à 16 ou moins, retirer le `convosync=direct` Option de montage car elle peut endommager les performances des E/S séquentielles. Cette étape nuit à d'autres aspects de la performance et ne doit être prise que si la valeur de `db_file_multiblock_read_count` doit être modifié par rapport à la valeur par défaut.

Linux

Rubriques de configuration spécifiques au système d'exploitation Linux.

Tables d'emplacements TCP Linux NFSv3

Les tables d'emplacements TCP sont l'équivalent NFSv3 de la profondeur de file d'attente de l'adaptateur de bus hôte (HBA). Ces tableaux contrôlent le nombre d'opérations NFS qui peuvent être en attente à la fois. La valeur par défaut est généralement 16, un chiffre bien trop faible pour assurer des performances optimales. Le problème inverse se produit sur les noyaux Linux plus récents : la limite de la table des emplacements TCP augmente automatiquement par envoi de demandes, jusqu'à atteindre le niveau de saturation du serveur NFS.

Pour des performances optimales et pour éviter les problèmes de performances, ajustez les paramètres du noyau qui contrôlent les tables d'emplacements TCP.

Exécutez le `sysctl -a | grep tcp.*.slot_table` et observez les paramètres suivants :

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tous les systèmes Linux doivent inclure `sunrpc.tcp_slot_table_entries`, mais seulement certains incluent `sunrpc.tcp_max_slot_table_entries`. Ils doivent tous deux être réglés sur 128.

> Si vous ne définissez pas ces paramètres, vous risquez d'avoir des effets importants sur les performances. Dans certains cas, les performances sont limitées car le système d'exploitation linux n'émet pas suffisamment d'E/S. Dans d'autres cas, les latences d'E/S augmentent à mesure que le système d'exploitation linux tente d'émettre plus d'E/S que ce qui peut être traité.

Options de montage NFS Linux

Le tableau suivant répertorie les options de montage NFS Linux pour une seule instance.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr`
ORACLE_HOME	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr`

Le tableau suivant répertorie les options de montage NFS Linux pour RAC.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, actimeo=0`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,actimeo=0`
CRS/vote	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,actimeo=0`
Ressource dédiée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144`
Partagée `ORACLE_HOME`	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,actimeo=0`

La principale différence entre les options de montage à instance unique et RAC est l'ajout de `actimeo=0` aux options de montage. Cet ajout a pour effet de désactiver la mise en cache du système d'exploitation hôte, ce qui permet à toutes les instances du cluster RAC d'avoir une vue cohérente de l'état des données. En utilisant le `init.ora` paramètre `filesystemio_options=setall` a le même effet que la désactivation de la mise en cache de l'hôte, il est toujours nécessaire de l'utiliser `actimeo=0`.

La raison `actimeo=0` est requis pour le partage `ORACLE_HOME` Les déploiements visent à faciliter la cohérence des fichiers tels que les fichiers de mots de passe Oracle et les fichiers spfiles. Si chaque instance d'un cluster RAC possède un dédié `ORACLE_HOME`, ce paramètre n'est pas requis.

En règle générale, les fichiers ne provenant pas de bases de données doivent être montés avec les mêmes options que celles utilisées pour les fichiers de données à instance unique. Toutefois, certaines applications peuvent avoir des exigences différentes. Évitez les options de montage `noac` et `actimeo=0` si possible parce que ces options désactivent la lecture et la mise en mémoire tampon au niveau du système de fichiers. Cela peut entraîner de graves problèmes de performances pour les processus tels que l'extraction, la translation et le chargement.

ACCESS et GETATTR

Certains clients ont remarqué qu'un niveau extrêmement élevé d'autres IOPS, comme L'ACCÈS et GETATTR, peut dominer leurs charges de travail. Dans des cas extrêmes, les opérations telles que les lectures et les écritures peuvent représenter jusqu'à 10 % du total. Il s'agit d'un comportement normal avec toute base de données qui inclut l'utilisation de `actimeo=0` et/ou `noac` Sous Linux car ces options font que le système d'exploitation Linux recharge en permanence les métadonnées de fichiers à partir du système de stockage. Les opérations telles que ACCESS et GETATTR sont des opérations à faible impact qui sont traitées à partir du cache ONTAP dans un environnement de base de données. Elles ne doivent pas être considérées comme des IOPS authentiques, comme les lectures et les écritures, qui génèrent une véritable demande pour les systèmes de stockage. Cependant, ces autres IOPS créent une certaine charge, en particulier dans les environnements RAC. Pour résoudre ce problème, activez dNFS, qui contourne le cache du tampon du système d'exploitation et évite ces opérations de métadonnées inutiles.

NFS direct Linux

Une option de montage supplémentaire, appelée `nosharecache`, Est requis lorsque (a) dNFS est activé et (b) qu'un volume source est monté plusieurs fois sur un seul serveur (c) avec un montage NFS imbriqué. Cette configuration est principalement utilisée dans les environnements prenant en charge les applications SAP. Par exemple, un seul volume sur un système NetApp peut avoir un répertoire situé sur `/vol/oracle/base` et une seconde à `/vol/oracle/home`. Si `/vol/oracle/base` est monté à `/oracle` et `/vol/oracle/home` est monté à `/oracle/home`, Le résultat est des montages NFS imbriqués qui proviennent de la même source.

Le système d'exploitation peut détecter le fait que `/oracle` et `/oracle/home` résident sur le même volume, qui est le même système de fichiers source. Le système d'exploitation utilise ensuite le même descripteur de périphérique pour accéder aux données. Cela améliore l'utilisation de la mise en cache du système d'exploitation et de certaines autres opérations, mais interfère avec dNFS. Si dNFS doit accéder à un fichier, par exemple `spfile` , sur , `/oracle/home` il peut essayer par erreur d'utiliser le mauvais chemin d'accès aux données. Le résultat est une opération d'E/S défaillante. Dans ces configurations, ajoutez `nosharecache` l'option de montage à tout système de fichiers NFS qui partage un volume source avec un autre système de fichiers NFS de cet hôte. Cela force le système d'exploitation Linux à allouer un descripteur de périphérique indépendant pour ce système de fichiers.

Linux Direct NFS et Oracle RAC

DNFS présente des avantages spéciaux en matière de performances pour Oracle RAC sur le système d'exploitation Linux. En effet, Linux ne dispose pas d'une méthode permettant de forcer les E/S directes, qui est requise avec RAC pour assurer la cohérence entre les nœuds. Pour contourner ce problème, Linux nécessite l'utilisation du `actimeo=0` Mount option, qui entraîne l'expiration immédiate des données de fichier à partir du cache du système d'exploitation. Cette option force à son tour le client Linux NFS à relire en permanence les données d'attributs, ce qui endommage la latence et augmente la charge sur le contrôleur de stockage.

L'activation de dNFS contourne le client NFS hôte et évite ces dommages. Plusieurs clients ont signalé une amélioration significative des performances sur les clusters RAC et une baisse significative de la charge ONTAP (en particulier par rapport aux autres IOPS) lors de l'activation de dNFS.

Linux Direct NFS et fichier orangfstab

Si vous utilisez dNFS sur Linux avec l'option de chemins d'accès multiples, vous devez utiliser plusieurs sous-réseaux. Sur d'autres systèmes d'exploitation, vous pouvez établir plusieurs canaux dNFS à l'aide du `LOCAL` et `DONTROUTE` Options de configuration de plusieurs canaux dNFS sur un même sous-réseau. Cependant, cela ne fonctionne pas correctement sur Linux et des problèmes de performances inattendus peuvent survenir. Sous Linux, chaque carte réseau utilisée pour le trafic dNFS doit se trouver sur un sous-réseau différent.

Planificateur d'E/S.

Le noyau Linux permet un contrôle de bas niveau sur la façon dont les E/S sont planifiées pour bloquer les périphériques. Les valeurs par défaut sur les différentes distributions de Linux varient considérablement. Les tests montrent que la date limite offre habituellement les meilleurs résultats, mais il arrive que le NOOP ait été légèrement meilleur. La différence de performance est minime, mais testez les deux options s'il est nécessaire d'extraire les performances maximales d'une configuration de base de données. Dans de nombreuses configurations, le paramètre CFQ est le paramètre par défaut. Il a démontré des problèmes de performances significatifs avec les charges de travail de la base de données.

Pour plus d'informations sur la configuration du planificateur d'E/S, reportez-vous à la documentation du fournisseur Linux correspondant.

Chemins d'accès multiples

Certains clients ont rencontré des pannes durant une interruption du réseau, car le démon multivoie ne s'exécutait pas sur leur système. Sur les versions récentes de Linux, le processus d'installation du système d'exploitation et le démon de chemins d'accès multiples peuvent exposer ces systèmes d'exploitation à ce problème. Les packages sont installés correctement, mais ils ne sont pas configurés pour un démarrage automatique après un redémarrage.

Par exemple, la valeur par défaut du démon multiacheminement sur RHEL5.5 peut apparaître comme suit :

```
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:off   3:off   4:off   5:off   6:off
```

Ceci peut être corrigé à l'aide des commandes suivantes :

```
[root@host1 iscsi]# chkconfig multipathd on
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off   1:off   2:on    3:on    4:on    5:on    6:off
```

Mise en miroir ASM

La mise en miroir ASM peut nécessiter des modifications des paramètres de chemins d'accès multiples Linux pour permettre à ASM de reconnaître un problème et de basculer vers un autre groupe de pannes. La plupart des configurations ASM sur ONTAP reposent sur une redondance externe. La protection des données est assurée par la baie externe et ASM ne met pas en miroir les données. Certains sites utilisent ASM avec redondance normale pour fournir une mise en miroir bidirectionnelle, généralement entre différents sites.

Les paramètres Linux indiqués dans le "Documentation des utilitaires hôtes NetApp" Incluez les paramètres de chemins d'accès multiples qui entraînent une mise en file d'attente illimitée des E/S. Cela signifie qu'une E/S sur un périphérique LUN sans chemin d'accès actif attend tant que les E/S sont terminées. Cette opération est généralement souhaitable, car les hôtes Linux attendent tant que nécessaire la fin des modifications du chemin SAN, le redémarrage des commutateurs FC ou le basculement d'un système de stockage.

Ce comportement de mise en file d'attente illimité cause un problème de mise en miroir ASM car ASM doit recevoir une erreur d'E/S pour qu'il puisse réessayer d'E/S sur une autre LUN.

Définissez les paramètres suivants dans Linux `multipath.conf` Fichier pour les LUN ASM utilisés avec la mise en miroir ASM :

```
polling_interval 5
no_path_retry 24
```

Ces paramètres créent une temporisation de 120 secondes pour les périphériques ASM. Le délai d'attente est calculé comme étant le `polling_interval` * `no_path_retry` en secondes. Il peut être nécessaire d'ajuster la valeur exacte dans certaines circonstances, mais un délai de 120 secondes doit être suffisant pour la plupart des utilisations. En particulier, 120 secondes doivent permettre un basculement ou un retour du contrôleur sans générer d'erreur d'E/S susceptible de mettre le groupe défaillant hors ligne.

Un plus bas `no_path_retry` La valeur peut réduire le temps nécessaire à ASM pour passer à un autre groupe de pannes, mais augmente également le risque de basculement indésirable lors des activités de maintenance, telles qu'une prise de contrôle. Le risque peut être atténué par une surveillance attentive de l'état de mise en miroir ASM. Si un basculement indésirable se produit, les miroirs peuvent être rapidement resynchronisés si la resynchronisation est effectuée relativement rapidement. Pour plus d'informations, consultez la documentation Oracle sur ASM Fast Mirror Resync pour la version du logiciel Oracle utilisé.

Options de montage Linux xfs, ext3 et ext4

NetApp recommande d'utiliser les options de montage par défaut.

ASMLib/AFD (pilote de filtre ASM)

Rubriques de configuration spécifiques au système d'exploitation Linux utilisant AFD et ASMlib

Tailles de bloc ASMlib

ASMlib est une bibliothèque de gestion ASM facultative et des utilitaires associés. Sa valeur principale est la capacité de tamponner un LUN ou un fichier NFS en tant que ressource ASM avec une étiquette lisible par l'utilisateur.

Les versions récentes d'ASMlib détectent un paramètre LUN appelé blocs logiques par exposant de bloc physique (LBPPBE). Cette valeur n'a été signalée que récemment par la cible SCSI ONTAP. Elle renvoie désormais une valeur qui indique qu'une taille de bloc de 4 Ko est recommandée. Il ne s'agit pas d'une définition de la taille de bloc, mais il est un indice pour toute application utilisant LBPPBE que les E/S d'une certaine taille peuvent être gérées plus efficacement. Cependant, ASMlib interprète LBPPBE comme une taille de bloc et estampille constamment l'en-tête ASM lors de la création du périphérique ASM.

Ce processus peut causer des problèmes avec les mises à niveau et les migrations de différentes manières, tous en fonction de l'incapacité à mélanger des périphériques ASMlib avec des tailles de bloc différentes dans le même groupe de disques ASM.

Par exemple, des tableaux plus anciens ont généralement signalé une valeur LBPPBE de 0 ou n'ont pas signalé cette valeur du tout. ASMlib l'interprète comme une taille de bloc de 512 octets. Pour les baies plus récentes, la taille de bloc est de 4 Ko. Il n'est pas possible de mélanger des périphériques de 512 octets et de 4 Ko dans le même groupe de disques ASM. Cela empêche un utilisateur d'augmenter la taille du groupe de disques ASM en utilisant des LUN de deux baies ou en utilisant ASM comme outil de migration. Dans d'autres cas, RMAN pourrait ne pas permettre la copie de fichiers entre un groupe de disques ASM avec une taille de bloc de 512 octets et un groupe de disques ASM avec une taille de bloc de 4 Ko.

La solution préférée est de corriger ASMlib. L'ID de bug Oracle est 13999609 et le correctif est présent dans

oracleasm-support-2.1.8-1 et versions ultérieures. Ce correctif permet à un utilisateur de définir le paramètre `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` à `true` dans le `/etc/sysconfig/oracleasm` fichier de configuration. Cela empêche ASMlib d'utiliser le paramètre LBPPBE, ce qui signifie que les LUN de la nouvelle baie sont maintenant reconnues comme des périphériques de bloc de 512 octets.

> L'option ne modifie pas la taille de bloc sur les LUN précédemment estampées par ASMlib. Par exemple, si un groupe de disques ASM avec des blocs de 512 octets doit être migré vers un nouveau système de stockage qui signale un bloc de 4 Ko, l'option `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` Doit être défini avant que les nouvelles LUN soient estampées avec ASMlib. Si les périphériques ont déjà été estampillés par oracleasm, ils doivent être reformatés avant d'être repoussées avec une nouvelle taille de bloc. Commencez par déconfigurer le périphérique avec `oracleasm deletedisk`, Puis effacez le premier 1 Go du périphérique avec `dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`. Enfin, si le périphérique a déjà été partitionné, utilisez le `kpartx` Commande permettant de supprimer les partitions obsolètes ou de simplement redémarrer le système d'exploitation.

Si ASMlib ne peut pas être corrigé, ASMlib peut être supprimé de la configuration. Ce changement est perturbateur et nécessite le démarquage des disques ASM et s'assurer que le `asm_diskstring` le paramètre est défini correctement. Toutefois, cette modification ne nécessite pas la migration des données.

Tailles de bloc d'entraînement de filtre ASM (AFD)

AFD est une bibliothèque de gestion ASM facultative qui remplace ASMlib. Du point de vue du stockage, il est très similaire à ASMlib, mais il inclut des fonctionnalités supplémentaires telles que la capacité de bloquer les E/S non-Oracle afin de réduire les risques d'erreurs d'utilisateur ou d'application susceptibles de corrompre les données.

Tailles des blocs de périphériques

Comme ASMlib, AFD lit également le paramètre LUN blocs logiques par exposant de bloc physique (LBPPBE) et utilise par défaut la taille de bloc physique, et non la taille de bloc logique.

Cela peut créer un problème si l'AFD est ajouté à une configuration existante où les périphériques ASM sont déjà formatés comme des périphériques de bloc de 512 octets. Le pilote AFD reconnaîtrait le LUN comme un périphérique 4K et l'incompatibilité entre l'étiquette ASM et le périphérique physique empêcherait l'accès. De même, les migrations seraient affectées, car il n'est pas possible de combiner des périphériques de 512 octets et de 4 Ko dans le même groupe de disques ASM. Cela empêche un utilisateur d'augmenter la taille du groupe de disques ASM en utilisant des LUN de deux baies ou en utilisant ASM comme outil de migration. Dans d'autres cas, RMAN pourrait ne pas permettre la copie de fichiers entre un groupe de disques ASM avec une taille de bloc de 512 octets et un groupe de disques ASM avec une taille de bloc de 4 Ko.

La solution est simple - AFD inclut un paramètre pour contrôler si elle utilise les tailles de bloc logiques ou physiques. Il s'agit d'un paramètre global affectant tous les périphériques du système. Pour forcer AFD à utiliser la taille de bloc logique, définissez `options oracleafd oracleafd_use_logical_block_size=1` dans le `/etc/modprobe.d/oracleafd.conf` fichier.

Tailles de transfert multivoie

Les modifications récentes du noyau linux appliquent des restrictions de taille d'E/S envoyées aux périphériques à chemins d'accès multiples, et AFD ne respecte pas ces restrictions. Les E/S sont ensuite rejetées, ce qui entraîne la mise hors ligne du chemin d'accès à la LUN. Il en résulte une incapacité à installer Oracle Grid, à configurer ASM ou à créer une base de données.

La solution consiste à spécifier manuellement la longueur de transfert maximale dans le fichier multipath.conf pour les LUN ONTAP :

```
devices {
            device {
                vendor "NETAPP"
                product "LUN.*"
                max_sectors_kb 4096
        }
    }
```

Même si aucun problème n'existe actuellement, ce paramètre doit être défini si l'AFD est utilisé pour garantir qu'une future mise à niveau de linux ne provoque pas de problèmes inattendus.

Microsoft Windows

Rubriques de configuration pour la base de données Oracle sous Microsoft Windows avec ONTAP.

NFS

Oracle prend en charge l'utilisation de Microsoft Windows avec le client NFS direct. Cette fonctionnalité offre les avantages de NFS en termes de gestion, notamment la possibilité d'afficher les fichiers dans les différents environnements, de redimensionner les volumes de façon dynamique et d'exploiter un protocole IP moins onéreux. Pour plus d'informations sur l'installation et la configuration d'une base de données sous Microsoft Windows à l'aide de dNFS, reportez-vous à la documentation officielle d'Oracle. Il n'existe pas de meilleures pratiques spéciales.

SAN

Pour une efficacité de compression optimale, assurez-vous que le système de fichiers NTFS utilise une unité d'allocation de 8 Ko ou plus. L'utilisation d'une unité d'allocation 4K, qui est généralement la valeur par défaut, a un impact négatif sur l'efficacité de la compression.

Solaris

Rubriques de configuration spécifiques au système d'exploitation Solaris.

Options de montage Solaris NFS

Le tableau suivant répertorie les options de montage Solaris NFS pour une seule instance.

Type de fichier	Options de montage
Accueil ADR	`rw,bg,hard,[vers=3,vers=4.1], roto=tcp, timeo=600,rsize=262144,wsize=262144`
Fichiers de contrôle Fichiers de données Journaux de reprise	`rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,llock,suid`

Type de fichier	Options de montage
ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid

L'utilisation de `llock` il a été prouvé qu'il améliorait considérablement les performances dans les environnements des clients en supprimant la latence associée à l'acquisition et au déblocage du système de stockage. Utilisez cette option avec soin dans les environnements dans lesquels de nombreux serveurs sont configurés pour monter les mêmes systèmes de fichiers et où Oracle est configuré pour monter ces bases de données. Bien qu'il s'agisse d'une configuration très inhabituelle, elle est utilisée par un petit nombre de clients. Si une instance est démarrée une seconde fois par erreur, une corruption des données peut se produire, car Oracle ne peut pas détecter les fichiers de verrouillage sur le serveur étranger. Les verrous NFS n'offrent pas de protection ; comme dans la version NFS 3, ils sont réservés à des conseils.

Parce que le `llock` et `forcedirectio` les paramètres s'excluent mutuellement, il est important que `filesystemio_options=setall` est présent dans le `init.ora` classez-les de sorte que `directio` est utilisé. Sans ce paramètre, la mise en cache du tampon du système d'exploitation hôte est utilisée et les performances peuvent être affectées.

Le tableau suivant répertorie les options de montage de Solaris NFS RAC.

Type de fichier	Options de montage
Accueil ADR	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, noac
Fichiers de contrôle Fichiers de données Journaux de reprise	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,forcedirectio
CRS/vote	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,forcedirectio
Ressource dédiée ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, suid
Partagée ORACLE_HOME	rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600,rsize=262144,wsize=262144, nointr,noac,suid

La principale différence entre les options de montage à instance unique et RAC est l'ajout de `noac` et `forcedirectio` aux options de montage. Cet ajout a pour effet de désactiver la mise en cache du système d'exploitation hôte, ce qui permet à toutes les instances du cluster RAC d'avoir une vue cohérente de l'état des données. En utilisant le `init.ora` paramètre `filesystemio_options=setall` a le même effet que la désactivation de la mise en cache de l'hôte, il est toujours nécessaire de l'utiliser `noac` et `forcedirectio`.

La raison `actimeo=0` est requis pour le partage `ORACLE_HOME` Les déploiements visent à faciliter la cohérence des fichiers tels que les fichiers de mots de passe Oracle et les fichiers spfiles. Si chaque instance d'un cluster RAC possède un dédié `ORACLE_HOME`, ce paramètre n'est pas requis.

Options de montage Solaris UFS

NetApp recommande fortement d'utiliser l'option de montage de journalisation afin de préserver l'intégrité des données en cas de panne de l'hôte Solaris ou d'interruption de la connectivité FC. L'option de montage de la journalisation préserve également l'utilisation des sauvegardes Snapshot.

ZFS Solaris

Solaris ZFS doit être installé et configuré avec soin pour offrir des performances optimales.

mvector

Solaris 11 a inclus un changement dans la façon dont il traite les opérations d'E/S importantes, ce qui peut entraîner de graves problèmes de performances sur les baies de stockage SAN. Le problème est documenté NetApp suivi bug report 630173, "Solaris 11 ZFS Performance regression".

Il ne s'agit pas d'un bogue de ONTAP. Il s'agit d'un défaut Solaris suivi sous les défauts Solaris 7199305 et 7082975.

Vous pouvez consulter le support Oracle pour savoir si votre version de Solaris 11 est affectée, ou vous pouvez tester la solution de contournement en la changeant `zfs_mvector_max_size` à une valeur plus petite.

Pour ce faire, exécutez la commande suivante en tant que root :

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t131072" |mdb -kw
```

En cas de problème inattendu résultant de cette modification, vous pouvez facilement l'inverser en exécutant la commande suivante en tant que root :

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t1048576" |mdb -kw
```

Noyau

Pour des performances ZFS fiables, un noyau Solaris est nécessaire pour résoudre les problèmes d'alignement des LUN. Le correctif a été introduit avec le correctif 147440-19 dans Solaris 10 et avec SRU 10.5 pour Solaris 11. Utilisez uniquement Solaris 10 et versions ultérieures avec ZFS.

Configuration du LUN

Pour configurer une LUN, effectuez les opérations suivantes :

1. Créer une LUN de type `solaris`.
2. Installez le kit d'utilitaire hôte (HUK) approprié spécifié par le "Matrice d'interopérabilité NetApp (IMT)".
3. Suivez les instructions du HUK exactement comme décrit. Les étapes de base sont décrites ci-dessous, mais reportez-vous au "documentation la plus récente" pour connaître la procédure adéquate.
   a. Exécutez le `host_config` utilitaire de mise à jour du `sd.conf/sdd.conf` fichier. Les disques SCSI seront ainsi en mesure de détecter correctement les LUN ONTAP.
   b. Suivez les instructions fournies par le `host_config` Utilitaire permettant d'activer les entrées/sorties

multivoies (MPIO).

c. Redémarrez. Cette étape est nécessaire pour que les modifications soient reconnues dans l'ensemble du système.

4. Partitionnez les LUN et vérifiez qu'ils sont correctement alignés. Voir « Annexe B : Vérification de l'alignement WAFL » pour obtenir des instructions sur la façon de tester et de confirmer directement l'alignement.

zpools

Un zpool ne doit être créé qu'après les étapes de la "Configuration du LUN" sont effectuées. Si la procédure n'est pas effectuée correctement, les performances risquent d'être sérieusement dégradées en raison de l'alignement des E/S. Pour des performances optimales sur ONTAP, les E/S doivent être alignées sur une limite de 4 Ko sur un disque. Les systèmes de fichiers créés sur un zpool utilisent une taille de bloc effective qui est contrôlée par un paramètre appelé `ashift`, qui peut être affiché en exécutant la commande `zdb -C`.

La valeur de `ashift` la valeur par défaut est 9, ce qui signifie 2^9, ou 512 octets. Pour des performances optimales, le `ashift` La valeur doit être 12 (2^12=4K). Cette valeur est définie au moment de la création du zpool et ne peut pas être modifiée, ce qui signifie que les données dans zpools avec `ashift` une migration autre que 12 doit être effectuée en copiant les données vers un nouveau zpool.

Après avoir créé un zpool, vérifiez la valeur de `ashift` avant de continuer. Si la valeur n'est pas 12, les LUN n'ont pas été détectées correctement. Détruisez le zpool, vérifiez que toutes les étapes indiquées dans la documentation des utilitaires hôtes correspondante ont été effectuées correctement et recréez le zpool.

Zpools et LDOMS Solaris

Les LDOMS Solaris créent une exigence supplémentaire pour s'assurer que l'alignement des E/S est correct. Bien qu'un LUN soit correctement découvert en tant que périphérique 4K, un périphérique virtuel vdsk sur un LDOM n'hérite pas de la configuration du domaine d'E/S. Le vdsk basé sur cette LUN revient par défaut à un bloc de 512 octets.

Un fichier de configuration supplémentaire est requis. Tout d'abord, les LDOM individuels doivent être corrigés pour le bogue Oracle 15824910 afin d'activer les options de configuration supplémentaires. Ce correctif a été porté dans toutes les versions actuellement utilisées de Solaris. Une fois le logiciel LDOM corrigé, il est prêt à configurer les nouveaux LUN correctement alignés comme suit :

1. Identifiez la ou les LUN à utiliser dans le nouveau zpool. Dans cet exemple, il s'agit du périphérique c2d1.

```
[root@LDOM1 ~]# echo | format
Searching for disks...done
AVAILABLE DISK SELECTIONS:
  0. c2d0 <Unknown-Unknown-0001-100.00GB>
     /virtual-devices@100/channel-devices@200/disk@0
  1. c2d1 <SUN-ZFS Storage 7330-1.0 cyl 1623 alt 2 hd 254 sec 254>
     /virtual-devices@100/channel-devices@200/disk@1
```

2. Récupérez l'instance vdc des systèmes à utiliser pour un pool ZFS :

```
[root@LDOM1 ~]#  cat /etc/path_to_inst
#
# Caution! This file contains critical kernel state
#
"/fcoe" 0 "fcoe"
"/iscsi" 0 "iscsi"
"/pseudo" 0 "pseudo"
"/scsi_vhci" 0 "scsi_vhci"
"/options" 0 "options"
"/virtual-devices@100" 0 "vnex"
"/virtual-devices@100/channel-devices@200" 0 "cnex"
"/virtual-devices@100/channel-devices@200/disk@0" 0 "vdc"
"/virtual-devices@100/channel-devices@200/pciv-communication@0" 0 "vpci"
"/virtual-devices@100/channel-devices@200/network@0" 0 "vnet"
"/virtual-devices@100/channel-devices@200/network@1" 1 "vnet"
"/virtual-devices@100/channel-devices@200/network@2" 2 "vnet"
"/virtual-devices@100/channel-devices@200/network@3" 3 "vnet"
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc" << We want
this one
```

3. Modifier `/platform/sun4v/kernel/drv/vdc.conf`:

```
block-size-list="1:4096";
```

Cela signifie que l'instance de périphérique 1 se voit attribuer une taille de bloc de 4096.

Par exemple, supposons que les instances vdsk 1 à 6 doivent être configurées pour une taille de bloc de 4 Ko et `/etc/path_to_inst` se lit comme suit :

```
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc"
"/virtual-devices@100/channel-devices@200/disk@2" 2 "vdc"
"/virtual-devices@100/channel-devices@200/disk@3" 3 "vdc"
"/virtual-devices@100/channel-devices@200/disk@4" 4 "vdc"
"/virtual-devices@100/channel-devices@200/disk@5" 5 "vdc"
"/virtual-devices@100/channel-devices@200/disk@6" 6 "vdc"
```

4. La finale `vdc.conf` le fichier doit contenir les éléments suivants :

```
block-size-list="1:8192","2:8192","3:8192","4:8192","5:8192","6:8192";
```

Avertissement
Le LDOM doit être redémarré après la configuration de vdc.conf et la création du vdsk. Cette étape ne peut pas être évitée. La modification de la taille de bloc n'est effective qu'après un redémarrage. Procéder à la configuration du pool de zpool et s'assurer que le module de transmission automatique est correctement réglé sur 12 comme décrit précédemment.

Journal des intentions ZFS (ZIL)

En général, il n'y a aucune raison de localiser le ZFS Intent Log (ZIL) sur un autre périphérique. Le journal peut partager de l'espace avec le pool principal. L'utilisation principale d'une ZIL distincte est l'utilisation de disques physiques qui n'offrent pas les fonctionnalités de mise en cache des écritures dans les baies de stockage modernes.

biais logique

Réglez le `logbias` Paramètre sur les systèmes de fichiers ZFS hébergeant les données Oracle.

```
zfs set logbias=throughput <filesystem>
```

Ce paramètre réduit les niveaux d'écriture globaux. Sous les valeurs par défaut, les données écrites sont d'abord validées dans le ZIL, puis dans le pool de stockage principal. Cette approche est adaptée à une configuration utilisant une configuration de disque simple, qui inclut un périphérique ZIL SSD et un support rotatif pour le pool de stockage principal. En effet, elle permet une validation dans une seule transaction d'E/S sur le support à latence la plus faible disponible.

Lorsque vous utilisez une baie de stockage moderne qui inclut sa propre capacité de mise en cache, cette approche n'est généralement pas nécessaire. Dans de rares cas, il peut être souhaitable d'effectuer une écriture avec une seule transaction dans le journal, par exemple une charge de travail composée d'écritures aléatoires hautement concentrées et sensibles à la latence. L'amplification d'écriture peut avoir des conséquences, car les données consignées sont finalement écrites dans le pool de stockage principal, ce qui double l'activité d'écriture.

E/S directes

De nombreuses applications, y compris les produits Oracle, peuvent contourner le cache du tampon hôte en activant des E/S directes Cette stratégie ne fonctionne pas comme prévu avec les systèmes de fichiers ZFS. Bien que le cache du tampon hôte soit contourné, ZFS lui-même continue à mettre en cache les données. Cette action peut entraîner des résultats trompeurs lors de l'utilisation d'outils tels que fio ou Sio pour effectuer des tests de performances. En effet, il est difficile de prévoir si les E/S atteignent le système de stockage ou si elles sont mises en cache localement au sein du système d'exploitation. Cette action rend également très difficile l'utilisation de tels tests synthétiques pour comparer les performances ZFS aux autres systèmes de fichiers. D'un point de vue pratique, les performances du système de fichiers varient considérablement, voire nulle, pour les charges de travail réelles des utilisateurs.

Plusieurs zpools

Les sauvegardes, les restaurations, les clones et l'archivage des données ZFS basés sur des snapshots doivent être effectués au niveau du zpool et requièrent généralement plusieurs zpools. Un zpool est similaire à un groupe de disques LVM et doit être configuré à l'aide des mêmes règles. Par exemple, il est probablement préférable de définir au mieux une base de données avec les fichiers de données résidant sur `zpool1` ainsi que les journaux d'archivage, les fichiers de contrôle et les journaux de reprise qui résident sur `zpool2`. Cette

approche permet une sauvegarde à chaud standard dans laquelle la base de données est placée en mode de sauvegarde à chaud, suivie d'un snapshot de `zpool1`. La base de données est alors supprimée du mode de sauvegarde à chaud, l'archivage des journaux est forcé et un instantané de `zpool2` est créé. Une opération de restauration nécessite de démonter les systèmes de fichiers zfs et de mettre hors ligne le zpool dans son intégralité, après une opération de restauration SnapRestore. Le zpool peut alors être remis en ligne et la base de données récupérée.

filesytemio_options

Le paramètre Oracle `filesystemio_options` Fonctionne différemment avec ZFS. Si `setall` ou `directio` Est utilisé, les opérations d'écriture sont synchrones et contournent le cache du tampon du système d'exploitation, mais les lectures sont mises en tampon par ZFS. Cette action engendre des difficultés dans l'analyse des performances, car les E/S sont parfois interceptées et traitées par le cache ZFS, ce qui rend la latence du stockage et les E/S totales inférieures à ce qu'elles semblent être.

Configuration du réseau

Interfaces logiques

Les bases de données Oracle doivent accéder au stockage. Les interfaces logiques (LIF) correspondent à la tuyauterie réseau qui connecte une machine virtuelle de stockage (SVM) au réseau et, par conséquent, à la base de données. Une conception correcte des LIF est requise pour s'assurer qu'il y a suffisamment de bande passante pour chaque charge de travail de la base de données, et le basculement ne provoque pas de perte des services de stockage.

Cette section présente les principes clés de conception des LIF. Pour obtenir une documentation plus complète, reportez-vous au "Documentation de gestion de réseau ONTAP". Comme pour les autres aspects de l'architecture de la base de données, les meilleures options pour la conception des machines virtuelles de stockage (SVM, appelé SVM au niveau de l'interface de ligne de commande) et de l'interface logique (LIF) dépendent largement des besoins en termes d'évolutivité et des besoins de l'entreprise.

Tenez compte des principaux sujets suivants lors de l'élaboration d'une stratégie LIF :

- **Performances.** la bande passante du réseau est-elle suffisante ?
- **Résilience.** y a-t-il des points de défaillance uniques dans la conception?
- **Gérabilité.** le réseau peut-il être mis à l'échelle sans interruption ?

Ces rubriques s'appliquent à la solution de bout en bout, de l'hôte aux commutateurs et au système de stockage.

Types de LIF

Il existe plusieurs types de LIF. "Documentation ONTAP sur les types de LIF" Fournir des informations plus complètes à ce sujet, mais d'un point de vue fonctionnel, les LIF peuvent être divisées en plusieurs groupes :

- **LIFs de gestion de clusters et de nœuds.** utilisées pour gérer le cluster de stockage.
- **LIF de gestion SVM.** interfaces permettant l'accès à une SVM via l'API REST ou ONTAPI (aussi connue sous le nom de ZAPI) pour des fonctions telles que la création de snapshots ou le redimensionnement de volumes. Des produits tels que SnapManager pour Oracle (SMO) doivent avoir accès à une LIF de gestion SVM.

- **Interfaces de données LIF.** pour FC, iSCSI, NVMe/FC, NVMe/TCP, NFS, ou SMB/CIFS.

> Une LIF de données utilisée pour le trafic NFS peut également être utilisée à des fins de gestion en modifiant la politique de pare-feu de `data` à `mgmt` Ou une autre règle autorisant HTTP, HTTPS ou SSH. Ce changement peut simplifier la configuration du réseau en évitant la configuration de chaque hôte pour l'accès à la fois à la LIF de données NFS et à une LIF de gestion distincte. Il n'est pas possible de configurer une interface pour l'iSCSI et le trafic de gestion, bien que les deux utilisent un protocole IP. Une LIF de gestion distincte est requise dans les environnements iSCSI.

Conception de SAN LIF

La conception de LIF dans un environnement SAN est relativement simple pour une raison : les chemins d'accès multiples. Toutes les implémentations SAN modernes permettent à un client d'accéder aux données sur plusieurs chemins réseau indépendants et de sélectionner le ou les chemins d'accès les plus adaptés. Par conséquent, les performances du design LIF sont plus simples à gérer, car les clients SAN équilibrent automatiquement la charge en E/S sur les meilleurs chemins disponibles.

Si un chemin devient indisponible, le client sélectionne automatiquement un autre chemin. La simplicité de conception qui en résulte rend les LIF SAN généralement plus faciles à gérer. Cela ne signifie pas pour autant qu'un environnement SAN est toujours plus facile à gérer, car de nombreux autres aspects du stockage SAN sont bien plus complexes que NFS. Cela signifie simplement que la conception de la LIF SAN est plus facile.

Performance

La bande passante est l'élément le plus important à prendre en compte dans les performances de LIF dans un environnement SAN. Par exemple, un cluster ONTAP AFF à deux nœuds doté de deux ports FC 16 Gb par nœud permet d'obtenir jusqu'à 32 Go de bande passante vers/depuis chaque nœud.

Résilience

Les LIF SAN ne basculent pas sur un système de stockage AFF. Si une LIF SAN échoue en raison du basculement du contrôleur, le logiciel de chemins d'accès multiples du client détecte la perte d'un chemin et redirige les E/S vers une autre LIF. Avec les systèmes de stockage ASA, les LIF basculent après un court délai, mais cela n'interrompt pas les E/S, car il existe déjà des chemins actifs sur l'autre contrôleur. Le processus de basculement a lieu afin de restaurer l'accès de l'hôte sur tous les ports définis.

Gestion aisée

La migration des LIF est une tâche beaucoup plus courante dans un environnement NFS, car elle est souvent associée au déplacement des volumes au sein du cluster. Il n'est pas nécessaire de migrer une LIF dans un environnement SAN lorsque les volumes sont déplacés au sein de la paire HA. En effet, une fois le déplacement de volume terminé, ONTAP envoie une notification au SAN concernant un changement de chemins et les clients SAN se réoptimisent automatiquement. La migration de LIF avec SAN est principalement associée à des modifications matérielles physiques majeures. Par exemple, si une mise à niveau des contrôleurs sans interruption est requise, une LIF SAN est migrée vers le nouveau matériel. Si un port FC est défectueux, une LIF peut être migrée vers un port non utilisé.

Recommandations de conception

NetApp fait les recommandations suivantes :

- Ne créez pas plus de chemins que nécessaire. Un nombre excessif de chemins complique la gestion globale et peut entraîner des problèmes de basculement de chemin sur certains hôtes. De plus, certains

hôtes ont des limites de chemin inattendues pour les configurations comme le démarrage SAN.

- Très peu de configurations doivent nécessiter plus de quatre chemins vers une LUN. L'intérêt d'avoir plus de deux nœuds de chemins publicitaires vers les LUN est limité, car l'agrégat hébergeant une LUN est inaccessible en cas de défaillance du nœud qui détient la LUN et de son partenaire haute disponibilité. Dans ce cas, la création de chemins sur des nœuds autres que la paire haute disponibilité principale n'est pas utile.
- Même si vous pouvez gérer le nombre de chemins de LUN visibles en sélectionnant les ports inclus dans les zones FC, il est généralement plus facile d'inclure tous les points cibles potentiels dans la zone FC et de contrôler la visibilité des LUN au niveau des ONTAP.
- Dans ONTAP 8.3 et versions ultérieures, la fonction de mappage de LUN sélectif (SLM) est la fonction par défaut. Avec SLM, toute nouvelle LUN est automatiquement annoncée à partir du nœud qui possède l'agrégat sous-jacent et du partenaire HA du nœud. Cet arrangement évite de créer des ensembles de ports ou de configurer le zoning pour limiter l'accessibilité des ports. Chaque LUN est disponible sur le nombre minimal de nœuds requis pour des performances et une résilience optimales.
  *Dans le cas où un LUN doit être migré en dehors des deux contrôleurs, les nœuds supplémentaires peuvent être ajoutés avec le `lun mapping add-reporting-nodes` De sorte que les LUN soient annoncées sur les nouveaux nœuds. Vous créez ainsi des chemins SAN supplémentaires vers les LUN pour la migration des LUN. Toutefois, l'hôte doit effectuer une opération de découverte pour utiliser les nouveaux chemins.
- Ne vous souciez pas trop du trafic indirect. Dans un environnement très exigeant en E/S, il est préférable d'éviter le trafic indirect pour lequel chaque microseconde de latence est critique, mais l'impact visible sur la performance est négligeable pour les charges de travail classiques.

Conception de LIF NFS

Contrairement aux protocoles SAN, NFS dispose d'une capacité limitée de définir plusieurs chemins d'accès aux données. Les extensions NFS parallèles (pNFS) à NFSv4 répondent à cette limitation, mais l'ajout de chemins d'accès supplémentaires devient rarement intéressant dans la mesure où les vitesses ethernet atteignent 100 Go et au-delà.

Performances et résilience

Bien que la mesure des performances d'une LIF SAN consiste principalement à calculer la bande passante totale à partir de tous les chemins principaux, la détermination des performances d'une LIF NFS nécessite d'étudier de plus près la configuration réseau exacte. Par exemple, deux ports 10 Gbit peuvent être configurés comme ports physiques bruts ou en tant que groupe d'interface LACP (Link Aggregation Control Protocol). S'ils sont configurés en tant que groupe d'interface, plusieurs stratégies d'équilibrage de charge sont disponibles et fonctionnent différemment selon que le trafic est commuté ou routé. Enfin, Oracle Direct NFS (dNFS) propose des configurations d'équilibrage de charge qui n'existent pour le moment dans aucun client OS NFS.

Contrairement aux protocoles SAN, les systèmes de fichiers NFS nécessitent une résilience au niveau de la couche de protocole. Par exemple, une LUN est toujours configurée avec les chemins d'accès multiples activés, ce qui signifie que plusieurs canaux redondants sont disponibles pour le système de stockage, chacun utilisant le protocole FC. Un système de fichiers NFS, en revanche, dépend de la disponibilité d'un seul canal TCP/IP qui ne peut être protégé qu'au niveau de la couche physique. C'est pourquoi des options telles que le basculement de port et l'agrégation de ports LACP existent.

Dans un environnement NFS, les performances et la résilience sont fournies au niveau de la couche du protocole réseau. En conséquence, ces deux sujets sont étroitement liés et doivent être discutés ensemble.

Lier les LIFs aux groupes de ports

Pour lier une LIF à un port group, associez l'adresse IP de la LIF à un groupe de ports physiques. La méthode principale pour agréger les ports physiques est le LACP. La fonctionnalité de tolérance aux pannes de LACP est assez simple : chaque port d'un groupe LACP est surveillé et supprimé du groupe de ports en cas de dysfonctionnement. Cependant, il existe de nombreuses idées fausses sur le fonctionnement de LACP en matière de performances :

- LACP ne requiert pas que la configuration sur le switch corresponde au terminal. Par exemple, ONTAP peut être configuré avec un équilibrage de charge basé sur IP, tandis qu'un commutateur peut utiliser un équilibrage de charge basé sur MAC.
- Chaque noeud final utilisant une connexion LACP peut choisir indépendamment le port de transmission des paquets, mais il ne peut pas choisir le port utilisé pour la réception. Cela signifie que le trafic de ONTAP vers une destination particulière est lié à un port particulier, et que le trafic de retour peut arriver sur une interface différente. Cela ne cause cependant aucun problème.
- LACP ne distribue pas uniformément le trafic en permanence. Dans un grand environnement comptant de nombreux clients NFS, le résultat est même généralement l'utilisation de tous les ports d'une agrégation LACP. Cependant, tout système de fichiers NFS dans l'environnement est limité à la bande passante d'un seul port, et non à l'agrégation complète.
- Bien que les politiques LACP robin-Robin soient disponibles sur ONTAP, ces règles n'abordent pas la connexion entre un switch et un hôte. Par exemple, une configuration avec une jonction LACP à quatre ports sur un hôte et une jonction LACP à quatre ports sur ONTAP ne peut toujours lire un système de fichiers qu'à l'aide d'un seul port. Bien que ONTAP puisse transmettre des données via les quatre ports, aucune technologie de commutation n'est actuellement disponible, qui envoie du commutateur à l'hôte via les quatre ports. Un seul est utilisé.

L'approche la plus courante dans les grands environnements composés de nombreux hôtes de base de données est de créer un agrégat LACP comportant un nombre approprié d'interfaces 10 Gbit (ou plus rapides) en utilisant l'équilibrage de la charge IP. Cette approche permet à ONTAP d'assurer une utilisation uniforme de tous les ports, tant qu'il y a suffisamment de clients. L'équilibrage de la charge est défaillant lorsque la configuration compte moins de clients, car les ressources en ligne LACP ne redistribuent pas la charge de manière dynamique.

Lorsqu'une connexion est établie, le trafic dans une direction particulière est placé sur un seul port. Par exemple, une base de données effectuant une analyse de table complète sur un système de fichiers NFS connecté via une jonction LACP à quatre ports lit les données via une seule carte d'interface réseau (NIC). Si seulement trois serveurs de base de données se trouvent dans un tel environnement, il est possible que les trois derniers lisent à partir du même port, alors que les trois autres ports sont inactifs.

Lier les LIF à des ports physiques

La liaison d'une LIF à un port physique permet un contrôle plus granulaire de la configuration du réseau, car une adresse IP donnée sur un système ONTAP n'est associée qu'à un seul port réseau à la fois. La résilience s'obtient ensuite via la configuration des groupes de basculement et des règles de basculement.

Stratégies de basculement et groupes de basculement

Le comportement des LIF durant une interruption du réseau est contrôlé par des règles de basculement et des groupes de basculement. Les options de configuration ont été modifiées avec les différentes versions de ONTAP. Consulter le "Documentation de gestion de réseau ONTAP pour les groupes et politiques de basculement" Pour plus d'informations sur la version de ONTAP déployée.

Les versions ONTAP 8.3 et supérieures permettent la gestion du basculement des LIF sur la base des

domaines de diffusion. Par conséquent, un administrateur peut définir tous les ports ayant accès à un sous-réseau donné et autoriser ONTAP à sélectionner une LIF de basculement appropriée. Cette approche peut être utilisée par certains clients, mais elle est limitée dans un environnement de réseau de stockage haut débit en raison du manque de prévisibilité. Par exemple, un environnement peut inclure à la fois des ports 1 Gbit pour l'accès aux systèmes de fichiers de routine et des ports 10 Gbit pour les E/S des fichiers de données Si les deux types de ports existent dans le même broadcast domain, le basculement de LIF peut entraîner le déplacement des E/S des fichiers de données d'un port 10 Gb vers un port 1 Gb.

En résumé, tenez compte des pratiques suivantes :

1. Configurez un groupe de basculement comme défini par l'utilisateur.
2. Remplissez le groupe de basculement avec les ports du contrôleur partenaire de basculement de stockage (SFO) de sorte que les LIF suivent les agrégats lors d'un basculement de stockage. Cela évite de créer du trafic indirect.
3. Utilisez les ports de basculement avec des caractéristiques de performance correspondantes à la LIF d'origine. Par exemple, une LIF située sur un seul port physique de 10 Go doit inclure un groupe de basculement doté d'un seul port 10 Go. Une LIF LACP à quatre ports doit basculer vers une autre LIF LACP à quatre ports. Ces ports seraient un sous-ensemble des ports définis dans le domaine de diffusion.
4. Définissez la politique de basculement sur partenaire SFO uniquement. Veillez donc à ce que la LIF suive l'agrégat lors du failover.

Restauration automatique

Réglez le `auto-revert` paramètre selon vos besoins. La plupart des clients préfèrent définir ce paramètre sur `true` Pour que la LIF rerevienne sur son port home. Cependant, dans certains cas, les clients ont défini cette option sur `false `afin qu'un basculement inattendu puisse être recherché avant de renvoyer une LIF à son port de attache.

Rapport LIF/volume

On croit souvent, à tort, qu'il doit y avoir une relation 1:1 entre les volumes et les LIFs NFS. Même si cette configuration est requise pour déplacer un volume n'importe où dans un cluster sans jamais créer de trafic d'interconnexion supplémentaire, elle n'est pas obligatoire de manière catégorique. Le trafic intercluster doit être envisagé, mais la simple présence du trafic intercluster ne crée pas de problèmes. Nombre des bancs d'essai publiés pour ONTAP portent sur des E/S principalement indirectes

Par exemple, un projet de base de données contenant un nombre relativement limité de bases de données pour lesquelles seuls 40 volumes nécessitent des performances élevées peut justifier un rapport volume 1:1 vers une stratégie LIF, un arrangement qui nécessiterait 40 adresses IP. N'importe quel volume peut ensuite être déplacé n'importe où dans le cluster avec la LIF associée, et le trafic serait toujours direct, minimisant ainsi chaque source de latence, même à des niveaux d'une microseconde.

Par exemple, un grand environnement hébergé peut être plus facilement géré avec une relation 1:1 entre les clients et les LIF. Au fil du temps, un volume peut avoir besoin d'être migré vers un autre nœud, ce qui provoque du trafic indirect. Cependant, l'effet sur les performances doit être indétectable à moins que les ports réseau du commutateur d'interconnexion ne soient saturés. En cas de problème, une nouvelle LIF peut être établie sur des nœuds supplémentaires et l'hôte peut être mis à jour dans la fenêtre de maintenance suivante afin de supprimer le trafic indirect de la configuration.

Configuration TCP/IP et ethernet

De nombreux clients d'Oracle sur ONTAP utilisent ethernet, le protocole réseau de NFS, iSCSI, NVMe/TCP, en particulier le cloud.

Paramètres du système d'exploitation hôte

La plupart des documents des fournisseurs d'applications incluent des paramètres TCP et ethernet spécifiques destinés à garantir le fonctionnement optimal de l'application. Ces mêmes paramètres suffisent généralement pour assurer des performances de stockage IP optimales.

Contrôle de flux Ethernet

Cette technologie permet à un client de demander à un expéditeur d'arrêter temporairement la transmission de données. Cela est généralement fait parce que le récepteur est incapable de traiter les données entrantes assez rapidement. À un moment donné, demander à un expéditeur de cesser la transmission était moins perturbant que d'avoir un récepteur de paquets de rejet parce que les tampons étaient pleins. Ce n'est plus le cas avec les piles TCP utilisées dans les systèmes d'exploitation d'aujourd'hui. En fait, le contrôle de flux cause plus de problèmes qu'il ne résout.

Les problèmes de performances causés par le contrôle de flux Ethernet ont augmenté ces dernières années. En effet, le contrôle de flux Ethernet fonctionne au niveau de la couche physique. Si une configuration réseau permet à un système d'exploitation hôte d'envoyer une demande de contrôle de flux Ethernet à un système de stockage, il en résulte une pause des E/S pour tous les clients connectés. Étant donné qu'un nombre croissant de clients sont servis par un seul contrôleur de stockage, la probabilité qu'un ou plusieurs de ces clients envoient des demandes de contrôle de flux augmente. Le problème a été fréquemment rencontré sur les sites des clients qui possèdent une virtualisation étendue du système d'exploitation.

Une carte réseau sur un système NetApp ne doit pas recevoir de demandes de contrôle de flux. La méthode utilisée pour obtenir ce résultat varie en fonction du fabricant du commutateur réseau. Dans la plupart des cas, le contrôle de flux sur un commutateur Ethernet peut être réglé sur `receive desired` ou `receive on`, ce qui signifie qu'une demande de contrôle de flux n'est pas transmise au contrôleur de stockage. Dans d'autres cas, la connexion réseau sur le contrôleur de stockage risque de ne pas permettre la désactivation du contrôle de flux. Dans ce cas, les clients doivent être configurés pour ne jamais envoyer de demandes de contrôle de flux, soit en changeant la configuration NIC sur le serveur hôte lui-même, soit en changeant les ports de commutateur auxquels le serveur hôte est connecté.

> **NetApp recommande** de s'assurer que les contrôleurs de stockage NetApp ne reçoivent pas de paquets de contrôle de flux Ethernet. Pour ce faire, il est généralement possible de définir les ports de commutateur auxquels le contrôleur est connecté, mais certains matériels de commutateur ont des limites qui peuvent nécessiter des modifications côté client.

Tailles du MTU

L'utilisation de trames Jumbo a été démontrée afin d'améliorer les performances des réseaux 1 Gbit en réduisant la surcharge du processeur et du réseau, mais l'avantage n'est généralement pas significatif.

> **NetApp recommande** d'implémenter des trames Jumbo lorsque cela est possible, à la fois pour réaliser des avantages potentiels en termes de performances et pour pérenniser la solution.

L'utilisation de trames Jumbo dans un réseau de 10 Gb est presque obligatoire. En effet, la plupart des implémentations de 10 Gbits atteignent une limite de paquets par seconde sans trames Jumbo avant d'atteindre le seuil de 10 Gbits. L'utilisation de trames jumbo améliore l'efficacité du traitement TCP/IP car elle permet au système d'exploitation, au serveur, aux cartes réseau et au système de stockage de traiter moins de paquets, mais des paquets plus volumineux. L'amélioration des performances varie d'une carte réseau à l'autre, mais elle est significative.

Dans le cas des implémentations de trames Jumbo, il est courant, mais incorrect, que tous les périphériques connectés doivent prendre en charge les trames Jumbo et que la taille MTU doit correspondre de bout en bout

Au lieu de cela, les deux extrémités du réseau négocient la taille de trame la plus élevée mutuellement acceptable lors de l'établissement d'une connexion. Dans un environnement standard, un commutateur réseau est défini sur une taille MTU de 9 9216, le contrôleur NetApp est défini sur 9000 et les clients sont configurés sur une combinaison de 9000 et 1514. Les clients qui prennent en charge un MTU de 9 9000 peuvent utiliser des trames jumbo, et les clients qui ne peuvent prendre en charge que 1514 peuvent négocier une valeur inférieure.

Les problèmes avec cet arrangement sont rares dans un environnement complètement commuté. Cependant, dans un environnement routé, veillez à ce qu'aucun routeur intermédiaire ne soit forcé de fragmenter des trames jumbo.

NetApp recommande de configurer les éléments suivants :

- Les trames Jumbo sont souhaitables, mais non requises avec Ethernet 1 Gb (GbE).
- Les trames Jumbo sont requises pour des performances maximales avec 10GbE et plus rapides.

Paramètres TCP

Trois paramètres sont souvent mal configurés : les horodatages TCP, l'acquittement sélectif (SACK) et la mise à l'échelle de la fenêtre TCP. De nombreux documents obsolètes sur Internet recommandent de désactiver un ou plusieurs de ces paramètres pour améliorer les performances. Cette recommandation a été très utile il y a de nombreuses années, lorsque les capacités du processeur étaient beaucoup plus faibles et qu'il y avait un avantage à réduire la surcharge sur le traitement TCP chaque fois que cela était possible.

Cependant, avec les systèmes d'exploitation modernes, la désactivation de l'une de ces fonctionnalités TCP n'entraîne généralement aucun avantage détectable, tout en pouvant nuire aux performances. Dans les environnements réseau virtualisés, les performances peuvent être endommagées, car ces fonctionnalités sont nécessaires pour gérer efficacement la perte de paquets et les modifications de la qualité du réseau.

NetApp recommande d'activer les horodatages TCP, le SACK et la mise à l'échelle des fenêtres TCP sur l'hôte, et ces trois paramètres doivent être activés par défaut dans tout système d'exploitation actuel.

Configuration FC SAN

La configuration de FC SAN pour les bases de données Oracle consiste principalement à suivre les meilleures pratiques quotidiennes en matière de SAN.

Il s'agit notamment de mesures de planification courantes, telles que l'assurance d'une bande passante suffisante sur le SAN entre l'hôte et le système de stockage, la vérification de la présence de tous les chemins SAN entre tous les périphériques requis, l'utilisation des paramètres de port FC requis par le fournisseur du commutateur FC, la prévention des conflits ISL, et à l'utilisation d'un système de surveillance de la structure SAN approprié.

Segmentation

Une zone FC ne doit jamais contenir plusieurs initiateurs. Un tel arrangement peut sembler fonctionner au départ, mais la diaphonie entre les initiateurs finit par interférer avec la performance et la stabilité.

Les zones à cibles multiples sont généralement considérées comme sûres, bien que dans de rares circonstances le comportement des ports cibles FC de fournisseurs différents ait causé des problèmes. Par exemple, évitez d'inclure les ports cibles d'une baie de stockage NetApp et non NetApp dans la même zone.

En outre, le fait de placer un système de stockage NetApp et un dispositif de bande dans la même zone est encore plus susceptible de causer des problèmes.

Réseau à connexion directe

Les administrateurs du stockage préfèrent parfois simplifier leurs infrastructures en supprimant les commutateurs réseau de la configuration. Cela peut être pris en charge dans certains scénarios.

ISCSI et NVMe/TCP

Un hôte utilisant iSCSI ou NVMe/TCP peut être directement connecté à un système de stockage et fonctionner normalement. La raison en est le chemin d'accès. Les connexions directes à deux contrôleurs de stockage distincts donnent lieu à deux chemins de flux de données indépendants. La perte du chemin, du port ou du contrôleur n'empêche pas l'autre chemin d'être utilisé.

NFS

Vous pouvez utiliser un stockage NFS à connexion directe, mais avec une limitation importante : le basculement ne fonctionnera pas sans script important, ce qui incombera au client.

Ce qui complique la reprise après incident avec un stockage NFS à connexion directe, c'est le routage qui se produit sur le système d'exploitation local. Par exemple, supposons qu'un hôte a une adresse IP 192.168.1.1/24 et qu'il est directement connecté à un contrôleur ONTAP avec une adresse IP 192.168.1.50/24. Lors du basculement, cette adresse 192.168.1.50 peut basculer vers l'autre contrôleur et sera disponible pour l'hôte, mais comment l'hôte peut-il détecter sa présence ? L'adresse 192.168.1.1 d'origine existe toujours sur la carte réseau hôte qui ne se connecte plus à un système opérationnel. Le trafic destiné à 192.168.1.50 continuerait d'être envoyé à un port réseau inutilisable.

Le second NIC du système d'exploitation peut être configuré sur 19 2.168.1.2 et serait capable de communiquer avec l'adresse en panne sur 192.168.1.50, mais les tables de routage locales auraient par défaut l'utilisation d'une adresse **et d'une seule adresse** pour communiquer avec le sous-réseau 192.168.1.0/24. Un administrateur système pourrait créer un framework de scripts qui détecterait une connexion réseau défaillante et modifierait les tables de routage locales ou rendrait les interfaces « up and down ». La procédure exacte dépend du système d'exploitation utilisé.

Dans la pratique, les clients NetApp disposent d'un protocole NFS à connexion directe, mais généralement uniquement pour les charges de travail où une pause des E/S est acceptable pendant les basculements. Lorsque des montages durs sont utilisés, aucune erreur d'E/S ne doit se produire lors de ces pauses. L'E/S doit se bloquer jusqu'à ce que les services soient restaurés, soit par un retour arrière, soit par une intervention manuelle pour déplacer les adresses IP entre les cartes réseau de l'hôte.

Connexion directe FC

Il n'est pas possible de connecter directement un hôte à un système de stockage ONTAP à l'aide du protocole FC. La raison en est l'utilisation de NPIV. Le WWN qui identifie un port FC ONTAP sur le réseau FC utilise un type de virtualisation appelé NPIV. Tout périphérique connecté à un système ONTAP doit pouvoir reconnaître un WWN NPIV. Aucun fournisseur actuel de HBA ne propose de HBA pouvant être installé sur un hôte et capable de prendre en charge une cible NPIV.

Configuration de stockage sous-jacente

SAN FC

Alignement de LUN

L'alignement des LUN fait référence à l'optimisation des E/S par rapport à la disposition du système de fichiers sous-jacent.

Directive non résolue dans <stdin> - include::lun-alignment.adoc[]

Voir également la discussion sur l'alignement des blocs de compression dans la section "Efficacité". Toute disposition alignée avec les limites des blocs de compression de 8 Ko est également alignée avec les limites de 4 Ko.

Avertissements de mauvais alignement

Directive non résolue dans <stdin> - include::database-alignment-warnings.adoc[]

L'alignement dans les environnements Solaris est plus compliqué. Reportez-vous à la section "Configuration de l'hôte SAN ONTAP" pour en savoir plus.

Avertissement
Dans les environnements Solaris x86, prenez davantage soin de l'alignement approprié car la plupart des configurations comportent plusieurs couches de partitions. Les tranches de partition Solaris x86 existent généralement au-dessus d'une table de partition d'enregistrement d'amorçage maître standard.

Dimensionnement des LUN et nombre de LUN

Il est essentiel de sélectionner la taille de LUN optimale et le nombre de LUN à utiliser pour optimiser les performances et la gestion des bases de données Oracle.

Une LUN est un objet virtualisé sur ONTAP qui existe sur tous les disques de l'agrégat d'hébergement. Par conséquent, les performances de la LUN ne sont pas affectées par sa taille, car la LUN exploite tout le potentiel de performance de l'agrégat, quelle que soit sa taille.

À titre de commodité, les clients peuvent souhaiter utiliser un LUN de taille spécifique. Par exemple, si une base de données est construite sur un groupe de disques LVM ou Oracle ASM composé de deux LUN de 1 To chacune, ce groupe de disques doit être développé par incréments de 1 To. Il peut être préférable de créer le groupe de disques à partir de huit LUN de 500 Go chacune, de sorte que le groupe de disques puisse être augmenté par incréments plus petits.

Il n'est pas recommandé d'établir une taille de LUN standard universelle, car cela peut compliquer la gestion. Par exemple, une taille de LUN standard de 100 Go peut fonctionner correctement lorsqu'une base de données ou un datastore se situe entre 1 et 2 To, mais qu'une base de données ou un datastore de 20 To nécessite 200 LUN. Cela signifie que les délais de redémarrage du serveur sont plus longs, que les différents utilisateurs doivent gérer davantage d'objets et que des produits tels que SnapCenter doivent effectuer des recherches sur de nombreux objets. L'utilisation d'un nombre inférieur de LUN de plus grande taille permet d'éviter de tels problèmes.

- Le nombre de LUN est plus important que la taille de LUN.
- La taille de LUN est principalement contrôlée par les exigences liées au nombre de LUN.
- Évitez de créer plus de LUN que nécessaire.

Nombre de LUN

Contrairement à la taille de LUN, le nombre de LUN affecte les performances. La performance des applications dépend souvent de la capacité à réaliser des E/S parallèles via la couche SCSI. Ainsi, deux LUN offrent de meilleures performances qu'une seule LUN. L'utilisation d'un LVM tel que Veritas VxVM, Linux LVM2 ou Oracle ASM est la méthode la plus simple pour augmenter le parallélisme.

Les clients NetApp n'ont généralement pas eu l'avantage d'augmenter le nombre de LUN au-delà de seize. Toutefois, le test d'environnements 100 % SSD avec des E/S aléatoires très lourdes a permis d'améliorer encore jusqu'à 64 LUN.

> **NetApp recommande** ce qui suit :
>
> En général, de quatre à seize LUN suffisent pour prendre en charge les besoins en E/S d'une charge de travail de base de données donnée. Moins de quatre LUN peuvent créer des limites de performances en raison de limites dans les implémentations SCSI hôte.

Placement des LUN

Le placement optimal des LUN de base de données dans les volumes ONTAP dépend principalement de l'utilisation des différentes fonctionnalités ONTAP.

Volumes

L'un des points de confusion les plus courants avec les nouveaux clients ONTAP est l'utilisation des volumes FlexVol, communément appelés simplement « volumes ».

Un volume n'est pas une LUN. Ces termes sont utilisés de façon synonymie avec de nombreux autres produits de fournisseurs, y compris les fournisseurs de cloud. Les volumes ONTAP sont des conteneurs de gestion simples. Ils ne fournissent pas les données en eux-mêmes, ni n'occupent l'espace. Il s'agit de conteneurs pour les fichiers et les LUN. Ils permettent d'améliorer et de simplifier la gestion, notamment à grande échelle.

Volumes et LUN

Les LUN associées sont généralement situées en colocation dans un seul volume. Par exemple, une base de données qui nécessite 10 LUN doit généralement avoir les 10 LUN placées sur le même volume.

> - L'utilisation d'un rapport LUN/volumes de 1:1, c'est-à-dire une LUN par volume, n'est **pas** une bonne pratique formelle.
> - À la place, les volumes doivent être considérés comme des conteneurs pour les charges de travail ou les datasets. Il peut y avoir une seule LUN par volume ou il peut y en avoir plusieurs. La bonne réponse dépend des exigences de gestion.
> - La diffusion des LUN sur un nombre inutile de volumes peut entraîner une surcharge supplémentaire et des problèmes de planification pour des opérations telles que les opérations de snapshot, un nombre excessif d'objets affichés dans l'interface utilisateur et entraîner l'atteinte des limites de volume de la plate-forme avant que la limite de LUN ne soit atteinte.

Volumes, LUN et snapshots

Les règles et planifications Snapshot sont placées sur le volume, et non sur la LUN. Un jeu de données composé de 10 LUN ne nécessite qu'une seule règle de snapshot lorsque ces LUN sont co-localisées dans le même volume.

En outre, la colocation de toutes les LUN associées à un jeu de données donné dans un seul volume permet d'effectuer des opérations de snapshot atomiques. Par exemple, une base de données résidant sur 10 LUN ou un environnement d'application VMware comprenant 10 systèmes d'exploitation différents peut être protégé comme un objet unique et cohérent si les LUN sous-jacentes sont tous placés sur un seul volume. S'ils sont placés sur des volumes différents, les snapshots peuvent être synchronisés à 100 %, même s'ils sont programmés en même temps.

Dans certains cas, il peut être nécessaire de diviser un jeu de LUN associé en deux volumes différents en raison des exigences de restauration. Par exemple, une base de données peut contenir quatre LUN pour les fichiers de données et deux LUN pour les journaux. Dans ce cas, un volume de fichiers de données avec 4 LUN et un volume de journaux avec 2 LUN peuvent être la meilleure option. La raison en est une capacité de restauration indépendante. Par exemple, le volume des fichiers de données peut être restauré de manière sélective à un état antérieur, ce qui signifie que les quatre LUN seraient rétablies à l'état du snapshot, tandis que le volume du journal contenant ses données stratégiques ne serait pas affecté.

Volumes, LUN et SnapMirror

Les règles et opérations SnapMirror sont, tout comme les opérations Snapshot, exécutées sur le volume, et non sur la LUN.

La colocation de LUN associées dans un seul volume vous permet de créer une relation SnapMirror unique et de mettre à jour toutes les données qu'elle contient en une seule mise à jour. Comme pour les instantanés, la mise à jour sera également une opération atomique. La destination SnapMirror dispose d'une réplique instantanée unique des LUN source. Si les LUN ont été réparties sur plusieurs volumes, les répliques peuvent être cohérentes les unes avec les autres.

Volumes, LUN et QoS

S'il est possible d'appliquer la QoS de manière sélective à chaque LUN, il est généralement plus facile de la configurer au niveau du volume. Par exemple, toutes les LUN utilisées par les invités dans un serveur ESX donné peuvent être placées sur un seul volume, puis une règle de qualité de service adaptative de ONTAP peut être appliquée. Vous obtenez ainsi une limite d'IOPS par To qui s'applique à toutes les LUN.

De même, si une base de données nécessitait 100 000 IOPS et occupait 10 LUN, il serait plus facile de définir une seule limite de 100 000 IOPS sur un seul volume que de définir 10 limites individuelles de 10 000 IOPS, une sur chaque LUN.

Dispositions multi-volumes

Dans certains cas, la distribution de LUN sur plusieurs volumes peut être avantageuse. La principale raison est la répartition des contrôleurs. Par exemple, un système de stockage haute disponibilité peut héberger une base de données unique dans laquelle chaque contrôleur a besoin du potentiel de traitement et de mise en cache complet. Dans ce cas, la conception type consisterait à placer la moitié des LUN dans un seul volume sur le contrôleur 1 et l'autre moitié des LUN dans un seul volume sur le contrôleur 2.

De même, la répartition des contrôleurs peut être utilisée pour l'équilibrage de la charge. Un système haute disponibilité hébergeant 100 bases de données de 10 LUN chacune peut être conçu où chaque base de données reçoit un volume de 5 LUN sur chacun des deux contrôleurs. Il en résulte une charge symétrique garantie de chaque contrôleur au fur et à mesure que des bases de données supplémentaires sont provisionnées.

Cependant, aucun de ces exemples ne correspond à un ratio volume/LUN de 1:1. L'objectif reste d'optimiser la gestion en co-localisant les LUN associées dans les volumes.

Par exemple, la conteneurisation est un rapport LUN/volume 1:1. Chaque LUN peut représenter une seule

charge de travail et doit être gérée individuellement. Dans ce cas, un rapport de 1:1 peut être optimal.

Redimensionnement des LUN et LVM

Lorsqu'un système de fichiers SAN a atteint sa limite de capacité, il existe deux options pour augmenter l'espace disponible :

- Augmentez la taille des LUN
- Ajoutez une LUN à un groupe de volumes existant et développez le volume logique contenu

Bien que le redimensionnement des LUN soit une option d'augmentation de la capacité, il est généralement préférable d'utiliser un LVM, y compris Oracle ASM. L'une des principales raisons pour lesquelles les LVM existent est d'éviter la nécessité d'un redimensionnement des LUN. Avec une LVM, plusieurs LUN sont reliées entre elles dans un pool de stockage virtuel. Les volumes logiques extraits de ce pool sont gérés par le LVM et peuvent être facilement redimensionnés. Il est également possible d'éviter les points sensibles sur un disque en distribuant un volume logique donné à tous les LUN disponibles. Une migration transparente peut généralement être effectuée à l'aide du gestionnaire de volumes pour déplacer les extensions sous-jacentes d'un volume logique vers de nouvelles LUN.

Répartition LVM

La répartition des LVM consiste à distribuer les données entre plusieurs LUN. Les performances de nombreuses bases de données en sont ainsi considérablement améliorées.

Avant l'ère des disques Flash, la répartition était utilisée pour surmonter les limites de performances des disques rotatifs. Par exemple, si un système d'exploitation doit effectuer une opération de lecture de 1 Mo, la lecture de ce 1 Mo de données à partir d'un seul disque demande beaucoup de tête de lecture lorsque le transfert des 1 Mo est lent. Si ce 1 Mo de données a été réparti sur 8 LUN, le système d'exploitation pourrait exécuter huit opérations de lecture de 128 K en parallèle et réduire le temps nécessaire au transfert de 1 Mo.

Le striping avec des disques rotatifs était plus difficile, car le modèle d'E/S devait être connu à l'avance. Si la répartition n'a pas été correctement réglée pour les véritables modèles d'E/S, les configurations à bandes risquent d'endommager les performances. Avec les bases de données Oracle, et en particulier les configurations 100 % Flash, le striping est beaucoup plus facile à configurer et a fait ses preuves pour améliorer considérablement les performances.

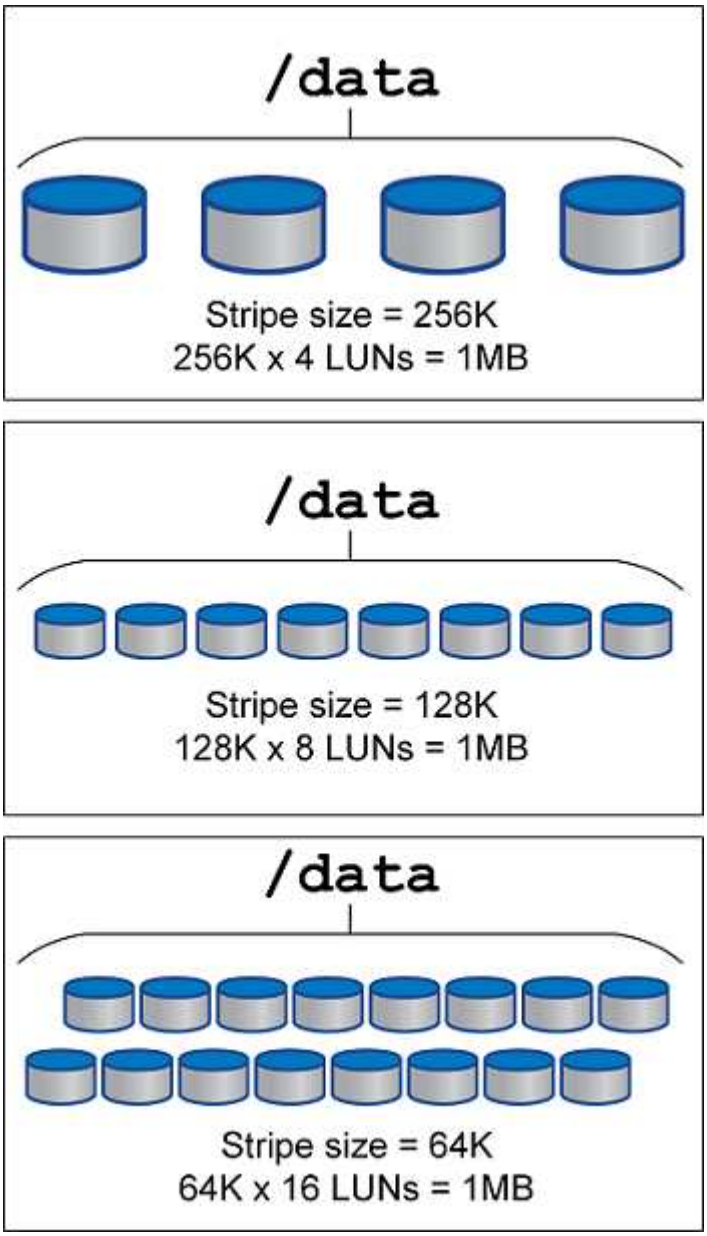
Par défaut, les gestionnaires de volumes logiques, tels que la bande Oracle ASM, ne le font pas pour le système d'exploitation natif LVM. Certaines lient plusieurs LUN ensemble en tant que périphérique concaténé. Résultat : des fichiers de données existent sur un seul périphérique LUN. Ceci provoque des points chauds. Les autres implémentations LVM prennent par défaut en charge les extensions distribuées. Cette méthode est similaire à la répartition, mais elle est plus grossière. Les LUN du groupe de volumes sont tranchées en grandes parties, appelées extensions et généralement mesurées en plusieurs mégaoctets. Ensuite, les volumes logiques sont distribués sur ces extensions. Il en résulte des E/S aléatoires sur un fichier qui doit être bien réparti entre les LUN, mais les opérations d'E/S séquentielles ne sont pas aussi efficaces qu'elles pourraient l'être.

Les E/S des applications exigeantes en performances sont presque toujours de (a) en unités de taille de bloc de base ou (b) d'un mégaoctet.

L'objectif principal d'une configuration à bandes est de s'assurer que les E/S de fichier unique peuvent être exécutées comme une seule unité, et que les E/S de plusieurs blocs, d'une taille de 1 Mo, peuvent être parallélisées de façon homogène sur toutes les LUN du volume réparti. Cela signifie que la taille de bande ne

doit pas être inférieure à la taille du bloc de base de données, et que la taille de bande multipliée par le nombre de LUN doit être de 1 Mo.

La figure suivante présente trois options possibles pour le réglage de la taille et de la largeur des bandes. Le nombre de LUN est sélectionné pour répondre aux exigences de performances comme décrit ci-dessus, mais dans tous les cas, le total des données dans une seule bande est de 1 Mo.



NFS

Présentation

NetApp fournit un stockage NFS haute performance depuis plus de 30 ans et son utilisation se développe avec les infrastructures basées sur le cloud en raison de sa simplicité.

Le protocole NFS comprend plusieurs versions aux exigences variables. Pour une description complète de la configuration NFS avec ONTAP, reportez-vous à la section "Tr-4067 NFS sur les meilleures pratiques ONTAP".

Les sections suivantes couvrent certaines des exigences les plus critiques et des erreurs utilisateur courantes.

Versions NFS

Le client NFS du système d'exploitation doit être pris en charge par NetApp.

- NFSv3 est pris en charge avec des systèmes d'exploitation conformes à la norme NFSv3.
- NFSv3 est pris en charge avec le client Oracle dNFS.
- NFSv4 est pris en charge avec tous les systèmes d'exploitation conformes à la norme NFSv4.
- NFSv4.1 et NFSv4.2 nécessitent une prise en charge spécifique du système d'exploitation. Consulter le "NetApp IMT" Pour les systèmes d'exploitation pris en charge.
- La prise en charge d'Oracle dNFS pour NFSv4.1 requiert Oracle 12.2.0.2 ou version supérieure.

> Le "Matrice de prise en charge de NetApp" Pour NFSv3 et NFSv4 n'incluent pas de systèmes d'exploitation spécifiques. Tous les systèmes d'exploitation conformes à la RFC sont généralement pris en charge. Lors d'une recherche dans la prise en charge en ligne de IMT pour NFSv3 ou NFSv4, ne sélectionnez pas de système d'exploitation spécifique, car aucune correspondance ne sera affichée. Tous les systèmes d'exploitation sont implicitement pris en charge par la politique générale.

Tables d'emplacements TCP Linux NFSv3

Les tables d'emplacements TCP sont l'équivalent NFSv3 de la profondeur de file d'attente de l'adaptateur de bus hôte (HBA). Ces tableaux contrôlent le nombre d'opérations NFS qui peuvent être en attente à la fois. La valeur par défaut est généralement 16, un chiffre bien trop faible pour assurer des performances optimales. Le problème inverse se produit sur les noyaux Linux plus récents : la limite de la table des emplacements TCP augmente automatiquement par envoi de demandes, jusqu'à atteindre le niveau de saturation du serveur NFS.

Pour des performances optimales et pour éviter les problèmes de performances, ajustez les paramètres du noyau qui contrôlent les tables d'emplacements TCP.

Exécutez le `sysctl -a | grep tcp.*.slot_table` et observez les paramètres suivants :

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tous les systèmes Linux doivent inclure `sunrpc.tcp_slot_table_entries`, mais seulement certains incluent `sunrpc.tcp_max_slot_table_entries`. Ils doivent tous deux être réglés sur 128.

> Si vous ne définissez pas ces paramètres, vous risquez d'avoir des effets importants sur les performances. Dans certains cas, les performances sont limitées car le système d'exploitation linux n'émet pas suffisamment d'E/S. Dans d'autres cas, les latences d'E/S augmentent à mesure que le système d'exploitation linux tente d'émettre plus d'E/S que ce qui peut être traité.

ADR et NFS

Certains clients ont signalé des problèmes de performances liés à une quantité excessive d'E/S dans le `ADR` emplacement. Le problème ne se produit généralement pas tant qu'une grande quantité de données de performances ne s'est pas accumulée. La raison de cet excès d'E/S est inconnue, mais ce problème semble

provenir des analyses répétées du répertoire cible par les processus Oracle pour détecter les modifications.

Dépose du `noac` et/ou `actimeo=0` Les options de montage permettent la mise en cache du système d'exploitation hôte et réduisent les niveaux d'E/S du stockage.

> **NetApp recommande** de ne pas placer `ADR` données sur un système de fichiers avec `noac` ou `actimeo=0` parce que des problèmes de performances sont probables. Séparer `ADR` le cas échéant, les données vers un autre point de montage.

nfs-rootonly et mount-rootonly

ONTAP inclut une option NFS appelée `nfs-rootonly` Cela permet de contrôler si le serveur accepte les connexions de trafic NFS à partir des ports élevés. Par mesure de sécurité, seul l'utilisateur root est autorisé à ouvrir des connexions TCP/IP à l'aide d'un port source inférieur à 1024 car ces ports sont normalement réservés à l'utilisation du système d'exploitation, et non aux processus utilisateur. Cette restriction permet de s'assurer que le trafic NFS provient d'un client NFS du système d'exploitation et non d'un processus malveillant émulant un client NFS. Le client Oracle dNFS est un pilote d'espace utilisateur, mais le processus s'exécute en tant que root, il n'est donc généralement pas nécessaire de modifier la valeur de `nfs-rootonly`. Les connexions sont réalisées à partir de ports bas.

Le `mount-rootonly` Cette option s'applique uniquement à NFSv3. Il contrôle si l'appel de MONTAGE RPC est accepté à partir de ports supérieurs à 1024. Lorsque dNFS est utilisé, le client est de nouveau exécuté en tant que root, ce qui lui permet d'ouvrir des ports inférieurs à 1024. Ce paramètre n'a aucun effet.

Les processus ouvrant des connexions avec dNFS sur les versions 4.0 et supérieures de NFS ne s'exécutent pas en tant que root et nécessitent donc des ports supérieurs à 1024. Le `nfs-rootonly` Le paramètre doit être défini sur Désactivé pour dNFS pour terminer la connexion.

Si `nfs-rootonly` Est activé, le résultat est un blocage lors de la phase de montage ouvrant les connexions dNFS. La sortie sqlplus ressemble à ceci :

```
SQL>startup
ORACLE instance started.
Total System Global Area 4294963272 bytes
Fixed Size                  8904776 bytes
Variable Size             822083584 bytes
Database Buffers         3456106496 bytes
Redo Buffers                7868416 bytes
```

Le paramètre peut être modifié comme suit :

```
Cluster01::> nfs server modify -nfs-rootonly disabled
```

> Dans de rares cas, vous devrez peut-être modifier nfs-rootonly et mount-rootonly sur Désactivé. Si un serveur gère un très grand nombre de connexions TCP, il est possible qu'aucun port inférieur à 1024 n'est disponible et que le système d'exploitation soit forcé d'utiliser des ports supérieurs. Ces deux paramètres ONTAP doivent être modifiés pour permettre la connexion.

Règles d'exportation NFS : superutilisateur et setuid

Si les binaires Oracle se trouvent sur un partage NFS, les règles d'export doivent inclure des autorisations de superutilisateur et de setuid.

Les exportations NFS partagées utilisées pour les services de fichiers génériques tels que les répertoires personnels des utilisateurs écraseront généralement l'utilisateur root. Cela signifie qu'une demande de l'utilisateur root sur un hôte qui a monté un système de fichiers est remappée en tant qu'utilisateur différent avec des privilèges inférieurs. Cela permet de sécuriser les données en empêchant un utilisateur root d'un serveur donné d'accéder aux données du serveur partagé. Le bit setuid peut également représenter un risque de sécurité dans un environnement partagé. Le bit setuid permet d'exécuter un processus en tant qu'utilisateur différent de celui qui appelle la commande. Par exemple, un script shell qui était détenu par root avec le bit setuid s'exécute en tant que root. Si ce script shell peut être modifié par d'autres utilisateurs, tout utilisateur non root peut émettre une commande en tant que root en mettant à jour le script.

Les binaires Oracle incluent les fichiers appartenant à root et utilisent le bit setuid. Si des binaires Oracle sont installés sur un partage NFS, les règles d'export doivent inclure les autorisations de superutilisateur et de setuid appropriées. Dans l'exemple ci-dessous, la règle inclut les deux `allow-suid` et permis `superuser` Accès (root) pour les clients NFS via l'authentification système.

```
Cluster01::> export-policy rule show -vserver vserver1 -policyname orabin
-fields allow-suid,superuser
vserver   policyname ruleindex superuser allow-suid
--------- ---------- --------- --------- ----------
vserver1  orabin     1         sys       true
```

Configuration NFSv4/4.1

Pour la plupart des applications, il y a très peu de différence entre NFSv3 et NFSv4. Les E/S applicatives sont généralement des E/S très simples et ne bénéficient pas énormément de certaines des fonctionnalités avancées de NFSv4. Les versions supérieures de NFS ne doivent pas être considérées comme une « mise à niveau » du point de vue du stockage de la base de données, mais plutôt comme des versions de NFS qui incluent des fonctionnalités supplémentaires. Par exemple, si la sécurité de bout en bout du mode de confidentialité kerberos (krb5p) est requise, NFSv4 est requis.

> **NetApp recommande** d'utiliser NFSv4.1 si les fonctionnalités NFSv4 sont requises. Certaines améliorations fonctionnelles du protocole NFSv4 dans NFSv4.1 améliorent la résilience dans certains cas à la périphérie.

Le passage à NFSv4 est plus compliqué que de simplement changer les options de montage de vers=3 en vers=4.1. Pour une explication plus complète de la configuration de NFSv4 avec ONTAP, notamment des conseils sur la configuration du système d'exploitation, voir "Tr-4067 NFS sur les meilleures pratiques ONTAP". Les sections suivantes de ce TR expliquent certaines des exigences de base relatives à l'utilisation de NFSv4.

Domaine NFSv4

Une explication complète de la configuration NFSv4/4.1 dépasse le cadre de ce document, mais un problème couramment rencontré est une incohérence dans le mappage de domaine. Du point de vue de sysadmin, les systèmes de fichiers NFS semblent se comporter normalement, mais les applications signalent des erreurs concernant les autorisations et/ou le setuid sur certains fichiers. Dans certains cas, les administrateurs ont conclu à tort que les autorisations des binaires de l'application ont été endommagées et ont exécuté des commandes chown ou chmod lorsque le problème réel était le nom de domaine.

Le nom de domaine NFSv4 est défini sur le SVM ONTAP :

```
Cluster01::> nfs server show -fields v4-id-domain
vserver   v4-id-domain
--------- ------------
vserver1  my.lab
```

Le nom de domaine NFSv4 sur l'hôte est défini dans `/etc/idmap.cfg`

```
[root@host1 etc]# head /etc/idmapd.conf
[General]
#Verbosity = 0
# The following should be set to the local NFSv4 domain name
# The default is the host's DNS domain name.
Domain = my.lab
```

Les noms de domaine doivent correspondre. Si ce n'est pas le cas, des erreurs de mappage similaires à ce qui suit apparaissent dans `/var/log/messages`:

```
Apr 12 11:43:08 host1 nfsidmap[16298]: nss_getpwnam: name 'root@my.lab'
does not map into domain 'default.com'
```

Les binaires d'application, tels que les binaires de base de données Oracle, incluent les fichiers appartenant à root avec le bit setuid, ce qui signifie qu'une discordance dans les noms de domaine NFSv4 provoque des échecs avec le démarrage d'Oracle et un avertissement sur la propriété ou les autorisations d'un fichier appelé `oradism`, qui est situé dans le `$ORACLE_HOME/bin` répertoire. Elle doit apparaître comme suit :

```
[root@host1 etc]# ls -l /orabin/product/19.3.0.0/dbhome_1/bin/oradism
-rwsr-x--- 1 root oinstall 147848 Apr 17  2019
/orabin/product/19.3.0.0/dbhome_1/bin/oradism
```

Si ce fichier apparaît avec la propriété de personne, il peut y avoir un problème de mappage de domaine NFSv4.

```
[root@host1 bin]# ls -l oradism
-rwsr-x--- 1 nobody oinstall 147848 Apr 17  2019 oradism
```

Pour résoudre ce problème, vérifiez le `/etc/idmap.cfg` Comparez le paramètre v4-ID-domain sur ONTAP et assurez-vous qu'ils sont cohérents. Si ce n'est pas le cas, effectuez les modifications requises, exécutez `nfsidmap -c`, et attendez un moment pour que les modifications se propagent. La propriété du fichier doit alors être correctement reconnue en tant que racine. Si un utilisateur a tenté de s'exécuter `chown root` Sur ce fichier avant que la configuration des domaines NFS ne soit corrigée, il peut être nécessaire de l'exécuter `chown root` encore.

Oracle Direct NFS (dNFS)

Les bases de données Oracle peuvent utiliser NFS de deux manières.

Tout d'abord, il peut utiliser un système de fichiers monté à l'aide du client NFS natif qui fait partie du système d'exploitation. Il s'agit parfois de kernel NFS ou KNFS. Le système de fichiers NFS est monté et utilisé par la base de données Oracle exactement comme toute autre application utiliserait un système de fichiers NFS.

La deuxième méthode est Oracle Direct NFS (dNFS). Il s'agit d'une implémentation de la norme NFS dans le logiciel de base de données Oracle. Elle ne modifie pas la façon dont les bases de données Oracle sont configurées ou gérées par l'administrateur de base de données. Tant que les paramètres du système de stockage lui-même sont corrects, l'utilisation de dNFS doit être transparente pour l'équipe DBA et les utilisateurs finaux.

Les systèmes de fichiers NFS habituels sont toujours montés sur une base de données avec la fonction dNFS activée. Une fois la base de données ouverte, la base de données Oracle ouvre un ensemble de sessions TCP/IP et effectue directement des opérations NFS.

NFS direct

La valeur principale de Direct NFS d'Oracle est de contourner le client NFS hôte et d'effectuer des opérations de fichiers NFS directement sur un serveur NFS. Pour l'activer, il suffit de modifier la bibliothèque Oracle Disk Manager (ODM). Vous trouverez des instructions sur ce processus dans la documentation Oracle.

L'utilisation de dNFS entraîne une amélioration significative des performances d'E/S et réduit la charge sur l'hôte et le système de stockage, car les E/S sont effectuées de la manière la plus efficace possible.

En outre, Oracle dNFS inclut une **option** pour les chemins d'accès multiples et la tolérance aux pannes de l'interface réseau. Par exemple, il est possible de lier deux interfaces de 10 Gbits pour offrir 20 Go de bande passante. En cas de défaillance d'une interface, les E/S sont relancées sur l'autre interface. L'opération globale est très similaire aux chemins d'accès multiples FC. Les chemins d'accès multiples étaient courants il y a plusieurs années, alors que l'ethernet 1 Gbit était la norme la plus courante. Une carte réseau 10 Go suffit pour la plupart des charges de travail Oracle, mais si un nombre supérieur de cartes réseau 10 Go sont requises, elles peuvent être reliées.

Lorsque dNFS est utilisé, il est essentiel que tous les correctifs décrits dans Oracle Doc 1495104.1 soient installés. Si un correctif ne peut pas être installé, l'environnement doit être évalué pour s'assurer que les bugs décrits dans ce document ne causent pas de problèmes. Dans certains cas, une incapacité à installer les correctifs requis empêche l'utilisation de dNFS.

N'utilisez pas dNFS avec tout type de résolution de noms round-Robin, y compris DNS, DDNS, NIS ou toute autre méthode. Cela inclut la fonction d'équilibrage de la charge DNS disponible dans ONTAP. Lorsqu'une base de données Oracle utilisant dNFS résout un nom d'hôte en adresse IP, elle ne doit pas être modifiée lors des recherches ultérieures. Cela peut entraîner des pannes de la base de données Oracle et une corruption potentielle des données.

Activation de dNFS

Oracle dNFS peut fonctionner avec NFSv3 sans aucune configuration nécessaire au-delà de l'activation de la bibliothèque dNFS (voir la documentation Oracle pour la commande spécifique requise). Toutefois, si dNFS ne parvient pas à établir la connectivité, il peut revenir en arrière silencieux au client NFS du noyau. Dans ce cas, les performances peuvent être gravement affectées.

Si vous souhaitez utiliser le multiplexage dNFS sur plusieurs interfaces, avec NFSv4.X, ou utiliser le chiffrement, vous devez configurer un fichier oranfstab. La syntaxe est extrêmement stricte. De petites erreurs

dans le fichier peuvent entraîner l'affichage du démarrage ou le contournement du fichier orangfstab.

Au moment de la rédaction de ce rapport, les chemins d'accès multiples dNFS ne fonctionnent pas avec NFSv4.1 avec les versions récentes d'Oracle Database. Un fichier oranfstab qui spécifie NFSv4.1 comme protocole ne peut utiliser qu'une instruction de chemin unique pour une exportation donnée. La raison en est que ONTAP ne prend pas en charge l'agrégation ClientID. Les correctifs de bases de données Oracle permettant de résoudre cette limitation seront peut-être disponibles à l'avenir.

La seule façon d'être certain que dNFS fonctionne comme prévu est d'interroger les tables v$dnfs.

Vous trouverez ci-dessous un exemple de fichier oranfstab situé dans /etc Il s'agit de l'un des emplacements multiples où un fichier oranfstab peut être placé.

```
[root@jfs11 trace]# cat /etc/oranfstab
server: NFSv3test
path: jfs_svmdr-nfs1
path: jfs_svmdr-nfs2
export: /dbf mount: /oradata
export: /logs mount: /logs
nfs_version: NFSv3
```

La première étape consiste à vérifier que dNFS est opérationnel pour les systèmes de fichiers spécifiés :

```
SQL> select dirname,nfsversion from v$dnfs_servers;

DIRNAME
------------------------------------
NFSVERSION
----------------
/logs
NFSv3.0

/dbf
NFSv3.0
```

Ce résultat indique que dNFS est utilisé avec ces deux systèmes de fichiers, mais que **pas** signifie que oranfstab est opérationnel. Si une erreur était présente, dNFS aurait détecté automatiquement les systèmes de fichiers NFS de l'hôte et il se peut que vous voyiez toujours la même sortie à partir de cette commande.

Les chemins d'accès multiples peuvent être vérifiés comme suit :

```
SQL> select svrname,path,ch_id from v$dnfs_channels;

SVRNAME
------------------------------------
PATH
------------------------------------
```

```
     CH_ID
----------
NFSv3test
jfs_svmdr-nfs1
         0

NFSv3test
jfs_svmdr-nfs2
         1

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------

NFSv3test
jfs_svmdr-nfs1
         0

NFSv3test
jfs_svmdr-nfs2

[output truncated]

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------
NFSv3test
jfs_svmdr-nfs2
         1

NFSv3test
jfs_svmdr-nfs1
         0

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------
```

```
NFSv3test
jfs_svmdr-nfs2
         1



66 rows selected.
```

Il s'agit des connexions que dNFS utilise. Deux chemins et canaux sont visibles pour chaque entrée SVRNAME. Cela signifie que les chemins d'accès multiples fonctionnent, ce qui signifie que le fichier oranfstab a été reconnu et traité.

Accès direct au NFS et au système de fichiers hôte

L'utilisation de dNFS peut parfois causer des problèmes pour les applications ou les activités des utilisateurs qui dépendent des systèmes de fichiers visibles montés sur l'hôte car le client dNFS accède au système de fichiers hors bande à partir du système d'exploitation hôte. Le client dNFS peut créer, supprimer et modifier des fichiers sans connaître le système d'exploitation.

Lorsque les options de montage des bases de données à instance unique sont utilisées, elles permettent la mise en cache des attributs de fichiers et de répertoires, ce qui signifie également que le contenu d'un répertoire est mis en cache. Par conséquent, dNFS peut créer un fichier, et il y a un court délai avant que le système d'exploitation ne relise le contenu du répertoire et que le fichier devienne visible pour l'utilisateur. Ce n'est généralement pas un problème, mais, dans de rares cas, des utilitaires tels que SAP BR*Tools peuvent présenter des problèmes. Si cela se produit, modifiez les options de montage pour utiliser les recommandations pour Oracle RAC. Ce changement entraîne la désactivation de l'ensemble de la mise en cache de l'hôte.

Ne modifiez les options de montage que si (a) dNFS est utilisé et (b) un problème résulte d'un décalage dans la visibilité des fichiers. Si dNFS n'est pas utilisé, les options de montage Oracle RAC sur une base de données à instance unique entraînent une dégradation des performances.

> Reportez-vous à la remarque à propos de `nosharecache` la "Options de montage NFS Linux" pour un problème dNFS spécifique à Linux qui peut produire des résultats inhabituels.

Locations et verrouillages NFS

NFSv3 est sans état. Cela signifie que le serveur NFS (ONTAP) ne suit pas les systèmes de fichiers montés, par qui ou quels verrous sont réellement en place.

ONTAP dispose de certaines fonctionnalités qui enregistreront les tentatives de montage. Vous savez donc quels clients accèdent aux données et il se peut que des verrous consultatifs soient présents, mais les informations ne sont pas 100 % complètes. Elle ne peut pas être terminée, car le suivi de l'état du client NFS ne fait pas partie de la norme NFSv3.

État NFSv4

En revanche, NFSv4 est avec état. Le serveur NFSv4 suit les clients qui utilisent les systèmes de fichiers, les fichiers existants, les fichiers et/ou les régions de fichiers verrouillés, etc Cela signifie qu'une communication régulière entre un serveur NFSv4 doit être établie pour maintenir les données d'état à jour.

Les États les plus importants gérés par le serveur NFS sont les verrous NFSv4 et les locations NFSv4, qui sont très étroitement liés. Vous devez comprendre comment chacun fonctionne par lui-même, et comment ils

se rapportent les uns aux autres.

Verrous NFSv4

Avec NFSv3, les verrous sont consultatifs. Un client NFS peut toujours modifier ou supprimer un fichier « verrouillé ». Un verrou NFSv3 n'expire pas de lui-même, il doit être supprimé. Cela crée des problèmes. Par exemple, si une application en cluster crée des verrous NFSv3 et que l'un des nœuds tombe en panne, que faire ? Vous pouvez coder l'application sur les nœuds survivants pour supprimer les verrous, mais comment savoir que c'est sûr ? Le nœud « en panne » est peut-être opérationnel, mais ne communique pas avec le reste du cluster ?

Avec NFSv4, les verrous ont une durée limitée. Tant que le client tenant les Locks continue à s'archiver avec le serveur NFSv4, aucun autre client n'est autorisé à acquérir ces Locks. Si un client ne parvient pas à s'archiver avec NFSv4, les verrous seront éventuellement révoqués par le serveur et d'autres clients pourront demander et obtenir des verrous.

Locations NFSv4

Les verrous NFSv4 sont associés à un bail NFSv4. Lorsqu'un client NFSv4 établit une connexion avec un serveur NFSv4, il obtient un bail. Si le client obtient un verrou (il existe plusieurs types de verrous), le verrou est associé au bail.

Ce bail a un délai défini. Par défaut, ONTAP définit la valeur de temporisation sur 30 secondes :

```
Cluster01::*> nfs server show -vserver vserver1 -fields v4-lease-seconds

vserver   v4-lease-seconds
--------- ----------------
vserver1  30
```

Cela signifie qu'un client NFSv4 doit vérifier avec le serveur NFSv4 toutes les 30 secondes pour renouveler ses baux.

Le bail est automatiquement renouvelé par n'importe quelle activité. Ainsi, si le client effectue des travaux, il n'est pas nécessaire d'effectuer des opérations supplémentaires. Si une application devient silencieuse et ne fait pas de véritable travail, elle devra effectuer une sorte d'opération de maintien en vie (appelée SÉQUENCE). Il s'agit essentiellement de dire « Je suis toujours là, veuillez actualiser mes contrats de location ».

```
 *Question:* What happens if you lose network connectivity for 31 seconds?
NFSv3 est sans état. Il ne s'attend pas à ce que les clients communiquent.
NFSv4 est avec état et une fois la période de location expirée, le bail
expire, et les verrous sont révoqués et les fichiers verrouillés sont mis
à disposition des autres clients.
```

Avec NFSv3, vous pouvez déplacer les câbles réseau, redémarrer les switchs réseau, modifier la configuration et être sûr qu'aucun problème ne se produirait. En général, les applications attendront patiemment le bon fonctionnement de la connexion réseau.

Avec NFSv4, vous disposez de 30 secondes (sauf si vous avez augmenté la valeur de ce paramètre dans

ONTAP) pour terminer votre travail. Si vous dépassez cette limite, vos contrats de location sont échus. Normalement, cela provoque des pannes d'application.

Par exemple, si vous disposez d'une base de données Oracle et que vous rencontrez une perte de connectivité réseau (parfois appelée « partition réseau ») qui dépasse le délai d'expiration du bail, vous plantez la base de données.

Voici un exemple de ce qui se passe dans le journal des alertes Oracle si cela se produit :

```
2022-10-11T15:52:55.206231-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00202: control file: '/redo0/NTAP/ctrl/control01.ctl'
ORA-27072: File I/O error
Linux-x86_64 Error: 5: Input/output error
Additional information: 4
Additional information: 1
Additional information: 4294967295
2022-10-11T15:52:59.842508-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00206: error in writing (block 3, # blocks 1) of control file
ORA-00202: control file: '/redo1/NTAP/ctrl/control02.ctl'
ORA-27061: waiting for async I/Os failed
```

Si vous examinez les syslog, vous devriez voir plusieurs de ces erreurs :

```
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
```

Les messages du journal sont généralement le premier signe d'un problème, autre que le blocage de l'application. En général, vous ne voyez rien pendant la panne réseau, car les processus et le système d'exploitation lui-même sont bloqués et tentent d'accéder au système de fichiers NFS.

Les erreurs apparaissent une fois que le réseau est de nouveau opérationnel. Dans l'exemple ci-dessus, une fois la connectivité rétablie, le système d'exploitation a tenté de réacquérir les verrous, mais il était trop tard. Le bail avait expiré et les serrures ont été retirées. Cela entraîne une erreur qui se propage jusqu'à la couche Oracle et provoque le message dans le journal des alertes. Vous pouvez voir des variations sur ces modèles en fonction de la version et de la configuration de la base de données.

En résumé, NFSv3 tolère l'interruption du réseau, mais NFSv4 est plus sensible et impose une période de location définie.

Que se passe-t-il si un délai de 30 secondes n'est pas acceptable ? Que se passe-t-il si vous gérez un réseau changeant de façon dynamique où les commutateurs sont redémarrés ou les câbles sont déplacés et que le résultat est une interruption occasionnelle du réseau ? Vous pouvez choisir de prolonger la période de

location, mais pour savoir si vous voulez y parvenir, vous devez expliquer les périodes de grâce NFSv4.

Périodes de grâce NFSv4

Lorsqu'un serveur NFSv3 est redémarré, il est prêt à transmettre les E/S presque instantanément. Il ne maintenait aucune sorte d'état concernant les clients. Le résultat est qu'une opération de basculement ONTAP semble souvent proche de l'instantané. Dès qu'un contrôleur est prêt à commencer à transmettre des données, il envoie un ARP au réseau qui signale le changement de topologie. En règle générale, les clients le détectent presque instantanément et le flux des données reprend.

NFSv4, cependant, fera une courte pause. Cela fait partie du fonctionnement de NFSv4.

> Les sections suivantes sont à jour depuis ONTAP 9.15.1, mais le comportement de bail et de verrouillage ainsi que les options de réglage peuvent changer de version à version. Si vous avez besoin d'ajuster les délais de location/verrouillage de NFSv4, veuillez consulter le support NetApp pour obtenir les informations les plus récentes.

Les serveurs NFSv4 doivent suivre les baux, les verrous et les utilisateurs des données. Si un serveur NFS fonctionne de manière incohérente et redémarre, ou perd de l'alimentation pendant un moment, ou est redémarré pendant l'activité de maintenance, le résultat est le bail/verrouillage et d'autres informations client sont perdues. Le serveur doit déterminer quel client utilise les données avant de reprendre les opérations. C'est là que intervient le délai de grâce.

Si vous mettez soudainement votre serveur NFSv4 hors/sous tension. Lorsqu'il est rétabli, les clients qui tentent de reprendre l'E/S reçoivent une réponse qui dit essentiellement « J'ai perdu les informations de location/verrouillage. Voulez-vous réenregistrer vos verrous ? » C'est le début de la période de grâce. La valeur par défaut est 45 secondes sur ONTAP :

```
Cluster01::> nfs server show -vserver vserver1 -fields v4-grace-seconds

vserver   v4-grace-seconds
--------- ----------------
vserver1  45
```

Par conséquent, après un redémarrage, un contrôleur met en pause les E/S tandis que tous les clients récupèrent leurs baux et verrous. Une fois le délai de grâce terminé, le serveur reprend les opérations d'E/S.

Cette période de grâce contrôle la récupération de bail pendant les modifications de l'interface réseau, mais il existe une deuxième période de grâce qui contrôle la récupération pendant le basculement du stockage `locking.grace_lease_seconds`. Il s'agit d'une option au niveau du nœud.

```
cluster01::> node run [node names or *] options
locking.grace_lease_seconds
```

Par exemple, si vous avez fréquemment besoin d'effectuer des basculements LIF, et que vous devez réduire le délai de grâce, vous changiez . `v4-grace-seconds` Si vous souhaitez améliorer le temps de reprise des E/S pendant le basculement du contrôleur, vous devez le modifier `locking.grace_lease_seconds`.

Ne modifiez ces valeurs qu'avec prudence et après avoir parfaitement compris les risques et les conséquences. Les pauses E/S liées aux opérations de basculement et de migration avec NFSv4.X ne

peuvent pas être entièrement évitées. Les périodes de verrouillage, de bail et de grâce font partie de la RFC NFS. Pour de nombreux clients, NFSv3 est préférable, car les délais de basculement sont plus courts.

Délais de location par rapport aux délais de grâce

Le délai de grâce et la période de location sont connectés. Comme mentionné ci-dessus, le délai de bail par défaut est de 30 secondes, ce qui signifie que les clients NFSv4 doivent s'enregistrer auprès du serveur au moins toutes les 30 secondes, sinon ils perdent leur bail et, à leur tour, leurs verrous. Le délai de grâce existe pour permettre à un serveur NFS de reconstruire les données de bail/verrouillage, et il prend par défaut 45 secondes. Le délai de grâce doit être plus long que la période de location. Cela permet de s'assurer qu'un environnement client NFS conçu pour renouveler les contrats de location au moins toutes les 30 secondes aura la possibilité d'archiver avec le serveur après un redémarrage. Un délai de grâce de 45 secondes garantit que tous les clients qui s'attendent à renouveler leur contrat de location au moins toutes les 30 secondes ont certainement l'occasion de le faire.

Si un délai de 30 secondes n'est pas acceptable, vous pouvez choisir de prolonger la période de location.

Si vous souhaitez augmenter le délai de bail à 60 secondes pour résister à une panne réseau de 60 secondes, vous devrez également augmenter le délai de grâce. Une pause d'E/S plus longue sera donc nécessaire lors du basculement du contrôleur.

Ce ne devrait normalement pas être un problème. En général, les utilisateurs ne mettent à jour les contrôleurs ONTAP qu'une ou deux fois par an. En outre, les basculements non planifiés en raison de défaillances matérielles sont extrêmement rares. En outre, si vous aviez un réseau où une panne réseau de 60 secondes était possible, et que le délai de bail était de 60 secondes, vous n'auriez probablement pas à vous opposer à un basculement rare du système de stockage, ce qui aurait entraîné une pause de 61 secondes non plus. Vous avez déjà reconnu que vous disposez d'un réseau qui s'arrête pendant plus de 60 secondes plutôt fréquemment.

Mise en cache NFS

La présence de l'une des options de montage suivantes entraîne la désactivation de la mise en cache de l'hôte :

```
cio, actimeo=0, noac, forcedirectio
```

Ces paramètres peuvent avoir un effet négatif important sur la vitesse d'installation du logiciel, de correction et des opérations de sauvegarde/restauration. Dans certains cas, en particulier avec les applications en cluster, ces options sont obligatoires car elles doivent inévitablement assurer la cohérence du cache sur tous les nœuds du cluster. Dans d'autres cas, les clients utilisent ces paramètres par erreur, ce qui entraîne des dommages inutiles aux performances.

De nombreux clients suppriment temporairement ces options de montage lors de l'installation ou de l'application de correctifs binaires. Cette suppression peut être effectuée en toute sécurité si l'utilisateur vérifie qu'aucun autre processus n'utilise activement le répertoire cible pendant le processus d'installation ou de correction.

Tailles de transfert NFS

Par défaut, ONTAP limite la taille des E/S NFS à 64 Ko.

Les E/S aléatoires utilisent la plupart des applications et bases de données une taille de bloc bien inférieure à la taille maximale de 64 Ko. Les E/S de blocs volumineux sont généralement parallélisées de sorte que le

maximum de 64 Ko ne limite pas non plus l'obtention d'une bande passante maximale.

Dans certains cas, le maximum de 64 000 charges de travail entraîne une limitation. En particulier, les opérations à thread unique, telles que les opérations de sauvegarde ou de restauration, ou encore les analyses de table complète de base de données s'exécutent plus rapidement et plus efficacement si la base de données peut exécuter moins d'E/S, mais plus volumineuses. La taille optimale de gestion des E/S pour ONTAP est de 256 Ko.

La taille maximale de transfert pour un SVM ONTAP donné peut être modifiée comme suit :

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

> Ne réduisez jamais la taille de transfert maximale autorisée sur ONTAP en dessous de la valeur de rsize/wsize des systèmes de fichiers NFS actuellement montés. Cela peut provoquer des blocages ou même une corruption des données avec certains systèmes d'exploitation. Par exemple, si les clients NFS sont actuellement définis sur une taille rsize/wsize de 65536, la taille maximale du transfert ONTAP peut être ajustée entre 65536 et 1048576 sans effet car les clients eux-mêmes sont limités. Réduire la taille de transfert maximale en dessous de 65536 peut endommager la disponibilité ou les données.

NVFAIL

NVFAIL est une fonctionnalité de ONTAP qui assure l'intégrité lors des scénarios de basculement catastrophiques.

En raison de la gestion de caches internes volumineux, les bases de données sont vulnérables à la corruption lors des événements de basculement du stockage. Si un événement catastrophique nécessite de forcer un basculement ONTAP ou de forcer le basculement MetroCluster, quel que soit l'état de santé de la configuration globale, les modifications qui ont été reconnues précédemment peuvent être supprimées. Le contenu de la matrice de stockage recule dans le temps et l'état du cache de la base de données ne reflète plus l'état des données sur le disque. Cette incohérence entraîne une corruption des données.

La mise en cache peut avoir lieu au niveau des applications ou des serveurs. Par exemple, une configuration Oracle Real application Cluster (RAC) avec des serveurs actifs sur un site principal et un site distant met en cache les données dans la SGA d'Oracle. Une opération de basculement forcé entraînant des pertes de données risque de corrompre la base de données, car les blocs stockés dans la mémoire SGA peuvent ne pas correspondre aux blocs du disque.

L'utilisation de la mise en cache est moins évidente au niveau du système de fichiers du système d'exploitation. Les blocs d'un système de fichiers NFS monté peuvent être mis en cache dans le système d'exploitation. Un système de fichiers en cluster basé sur des LUN situés sur le site principal peut également être monté sur des serveurs du site distant, et une fois encore, les données peuvent être mises en cache. Une défaillance de la mémoire NVRAM, un basculement forcé ou un basculement forcé dans ces situations peuvent entraîner une corruption du système de fichiers.

ONTAP protège les bases de données et les systèmes d'exploitation de ce scénario avec NVFAIL et ses paramètres associés.

Utilitaire de récupération ASM (ASMRU)

ONTAP supprime efficacement les blocs nuls écrits sur un fichier ou une LUN lorsque la compression à la volée est activée. Des utilitaires tels que l'utilitaire ASRU (Oracle ASM Reclamation Utility) sont utilisés en écrivant des zéros dans les extensions ASM inutilisées.

Cela permet aux administrateurs de bases de données de récupérer de l'espace sur la baie de stockage après la suppression des données. ONTAP intercepte les zéros et désalloue l'espace de la LUN. Le processus de récupération est extrêmement rapide, car aucune donnée n'est écrite dans le système de stockage.

Du point de vue de la base de données, le groupe de disques ASM contient des zéros et la lecture de ces régions des LUN entraîne un flux de zéros, mais ONTAP ne stocke pas les zéros sur les disques. Des modifications simples des métadonnées sont effectuées en interne pour marquer les régions mises à zéro de la LUN comme vides de toutes les données.

Pour des raisons similaires, le test de performance impliquant des données mises à zéro n'est pas valide, car les blocs de zéros ne sont pas réellement traités comme des écritures dans la baie de stockage.

> Lorsque vous utilisez ASRU, assurez-vous que tous les correctifs recommandés par Oracle sont installés.

Virtualisation

La virtualisation des bases de données avec VMware, Oracle OLVM ou KVM est un choix de plus en plus courant pour les clients NetApp qui ont choisi la virtualisation, même pour leurs bases de données les plus stratégiques.

Prise en charge

Il existe de nombreuses idées fausses sur les politiques de prise en charge d'Oracle pour la virtualisation, en particulier pour les produits VMware. Il n'est pas rare d'entendre qu'Oracle ne prend pas en charge la virtualisation. Cette notion est incorrecte et ne permet pas de bénéficier de la virtualisation. Oracle Doc ID 249212.1 traite des besoins réels et est rarement considéré par les clients comme un problème.

Si un problème se produit sur un serveur virtualisé et que ce problème n'est pas encore connu du support Oracle, le client peut être invité à reproduire le problème sur du matériel physique. Si un client Oracle exécute une version de pointe d'un produit, il est possible qu'il ne souhaite pas utiliser la virtualisation en raison de problèmes de prise en charge potentiels, mais cette situation n'est pas réelle pour les clients qui utilisent des versions de produits Oracle généralement disponibles.

Présentation du stockage

Les clients qui envisagent de virtualiser leurs bases de données doivent baser leurs décisions de stockage sur leurs besoins métier. Bien qu'il s'agisse généralement d'une véritable déclaration pour toutes les décisions IT, elle est particulièrement importante pour les projets de base de données, car la taille et le champ d'application des exigences varient considérablement.

Il existe trois options de base pour la présentation du stockage :

- LUN virtualisées sur les datastores de l'hyperviseur
- LUN iSCSI gérées par l'initiateur iSCSI sur la machine virtuelle, pas par l'hyperviseur
- Systèmes de fichiers NFS montés par la machine virtuelle (pas à partir d'un datastore basé sur NFS)
- Mappages directs de périphériques. Les clients ne préfèrent pas les RDM VMware, mais les périphériques physiques sont souvent mappés directement avec la virtualisation KVM et OLVM.

Performance

La méthode de présentation du stockage à un invité virtualisé n'a généralement pas d'incidence sur les performances. Les systèmes d'exploitation hôtes, les pilotes réseau virtualisés et les implémentations de datastores d'hyperviseurs sont tous optimisés et peuvent généralement utiliser toute la bande passante réseau FC ou IP disponible entre l'hyperviseur et le système de stockage, dans la mesure où les meilleures pratiques de base sont respectées. Dans certains cas, il peut être légèrement plus facile d'obtenir des performances optimales en utilisant une approche de présentation du stockage par rapport à une autre, mais le résultat final devrait être comparable.

Gestion aisée

Le facteur clé dans la décision de présenter le stockage à un invité virtualisé est la mangeabilité. Il n'y a pas de bonne ou de mauvaise méthode. La meilleure approche dépend des besoins, des compétences et des préférences opérationnels de l'IT.

Les facteurs à prendre en compte sont les suivants :

- **Transparence.** lorsqu'une machine virtuelle gère ses systèmes de fichiers, il est plus facile pour un administrateur de base de données ou un administrateur système d'identifier la source des systèmes de fichiers pour leurs données. L'accès aux systèmes de fichiers et aux LUN est différent de celui d'un serveur physique.
- **Cohérence.** lorsqu'une machine virtuelle possède ses systèmes de fichiers, l'utilisation ou la non-utilisation d'une couche hyperviseur affecte la gestion. Les mêmes procédures de provisionnement, de surveillance, de protection des données, etc. Peuvent être utilisées dans l'ensemble du parc, y compris dans les environnements virtualisés et non virtualisés.

  Par contre, dans un data Center 100 % virtualisé, il peut être préférable d'utiliser un stockage basé sur un datastore pour l'ensemble de l'encombrement, selon la même logique que celle mentionnée ci-dessus. Cohérence : possibilité d'utiliser les mêmes procédures de provisionnement, de protection, de regroupement et de protection des données.

- **Stabilité et dépannage.** lorsqu'une machine virtuelle possède ses systèmes de fichiers, il est plus simple de fournir des performances stables et de résoudre les problèmes car la pile de stockage complète est présente sur la machine virtuelle. Le seul rôle de l'hyperviseur est de transporter des trames FC ou IP. Lorsqu'un datastore est inclus dans une configuration, il complique la configuration en introduisant un autre ensemble d'expirations de délai, de paramètres, de fichiers journaux et de bogues potentiels.
- **Portabilité.** lorsqu'une machine virtuelle possède ses systèmes de fichiers, le processus de déplacement d'un environnement Oracle devient beaucoup plus simple. Les systèmes de fichiers peuvent facilement être déplacés entre des invités virtualisés et non virtualisés.
- **Dépendance vis-à-vis d'un fournisseur.** une fois les données placées dans un datastore, il devient difficile d'utiliser un hyperviseur différent ou de retirer les données de l'environnement virtualisé.
- **Activation de Snapshot.** les procédures de sauvegarde traditionnelles dans un environnement virtualisé peuvent devenir problématiques en raison de la bande passante relativement limitée. Par exemple, un

agrégat 10 GbE à quatre ports peut suffire pour répondre aux besoins quotidiens en performances de nombreuses bases de données virtualisées, mais ce trunk ne permet pas d'effectuer des sauvegardes à l'aide de RMAN ou d'autres produits de sauvegarde nécessitant le streaming d'une copie complète des données. Résultat : un environnement virtualisé de plus en plus consolidé doit effectuer des sauvegardes via des snapshots de stockage. Ainsi, il n'est pas nécessaire de surconstruire la configuration de l'hyperviseur uniquement pour prendre en charge les besoins en bande passante et en CPU dans la fenêtre de sauvegarde.

L'utilisation de systèmes de fichiers détenus par les clients facilite parfois l'exploitation des sauvegardes et des restaurations basées sur des snapshots, car les objets de stockage nécessitant une protection peuvent être ciblés plus facilement. Cependant, de plus en plus de produits de protection des données de virtualisation s'intègrent bien aux datastores et aux snapshots. La stratégie de sauvegarde doit être totalement adoptée avant de décider comment présenter le stockage à un hôte virtualisé.

Pilotes paravirtualisés

Pour des performances optimales, l'utilisation de pilotes de réseau paravirtualisés est essentielle. Lorsqu'un datastore est utilisé, un pilote SCSI paravirtualisé est requis. Un pilote de périphérique paravirtualisé permet à un invité de s'intégrer plus profondément dans l'hyperviseur, au lieu d'un pilote émulé dans lequel l'hyperviseur passe plus de temps CPU à imiter le comportement du matériel physique.

Saturation de la mémoire RAM

La saturation de la mémoire RAM implique la configuration d'une quantité de mémoire RAM virtualisée supérieure à celle qui existe sur le matériel physique sur différents hôtes. Cela peut entraîner des problèmes de performances inattendus. Lors de la virtualisation d'une base de données, les blocs sous-jacents de la SGA d'Oracle ne doivent pas être remplacés par l'hyperviseur vers le stockage. Cela entraîne des résultats de performances très instables.

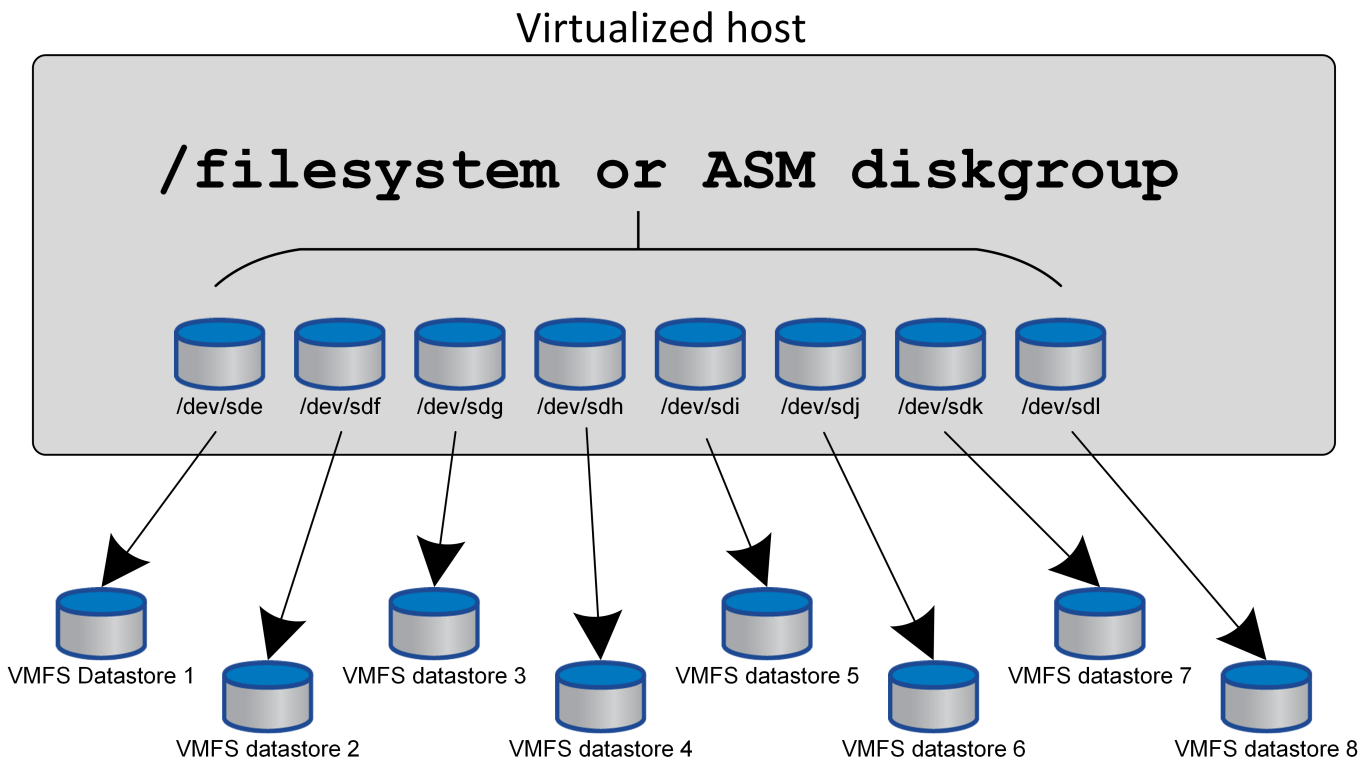
Répartition des datastores

Lorsque vous utilisez des bases de données avec des datastores, un facteur critique est à prendre en compte en ce qui concerne la répartition des performances.

Les technologies de datastore, telles que VMFS, peuvent couvrir plusieurs LUN, mais ne sont pas des périphériques répartis. Les LUN sont concaténées. Il peut en résulter des points sensibles de la LUN. Par exemple, une base de données Oracle standard peut disposer d'un groupe de disques ASM à 8 LUN. Les 8 LUN virtualisées pourraient être provisionnées sur un datastore VMFS de 8 LUN, mais il n'y a aucune garantie sur les LUN sur lesquelles les données résideront. La configuration résultante peut être les 8 LUN virtualisées occupant une seule LUN au sein du datastore VMFS. Cela risque d'engorgement des performances.

La répartition est généralement requise. Avec certains hyperviseurs, dont KVM, il est possible de créer un datastore à l'aide de la répartition LVM, comme décrit ci-dessous "ici". Avec VMware, l'architecture semble un peu différente. Chaque LUN virtualisée doit être placée sur un datastore VMFS différent.

Par exemple :



Le facteur principal de cette approche n'est pas le ONTAP. En raison de la limitation inhérente du nombre d'opérations qu'une seule machine virtuelle ou LUN d'hyperviseur peut traiter en parallèle, En règle générale, un LUN ONTAP peut prendre en charge beaucoup plus d'IOPS qu'un hôte ne peut en demander. La limite de performances d'une seule LUN est presque universellement due au système d'exploitation hôte. Ainsi, la plupart des bases de données ont besoin de 4 à 8 LUN pour répondre à leurs besoins de performance.

Les architectures VMware doivent planifier soigneusement leurs architectures pour s'assurer que cette approche ne permet pas d'optimiser le datastore et/ou le chemin LUN. Par ailleurs, il n'est pas nécessaire de disposer d'un ensemble unique de datastores VMFS pour chaque base de données. Le principal besoin est de s'assurer que chaque hôte dispose d'un ensemble propre de 4-8 chemins d'E/S entre les LUN virtualisées et les LUN back-end sur le système de stockage lui-même. Dans de rares cas, des exigences de performances vraiment extrêmes peuvent se révéler bénéfiques pour encore plus de données, mais 4-8 LUN suffisent généralement pour 95 % de toutes les bases de données. Un volume ONTAP unique contenant 8 LUN peut prendre en charge jusqu'à 250,000 000 IOPS de bloc Oracle aléatoires avec une configuration type OS/ONTAP/réseau.

Tiering

Présentation

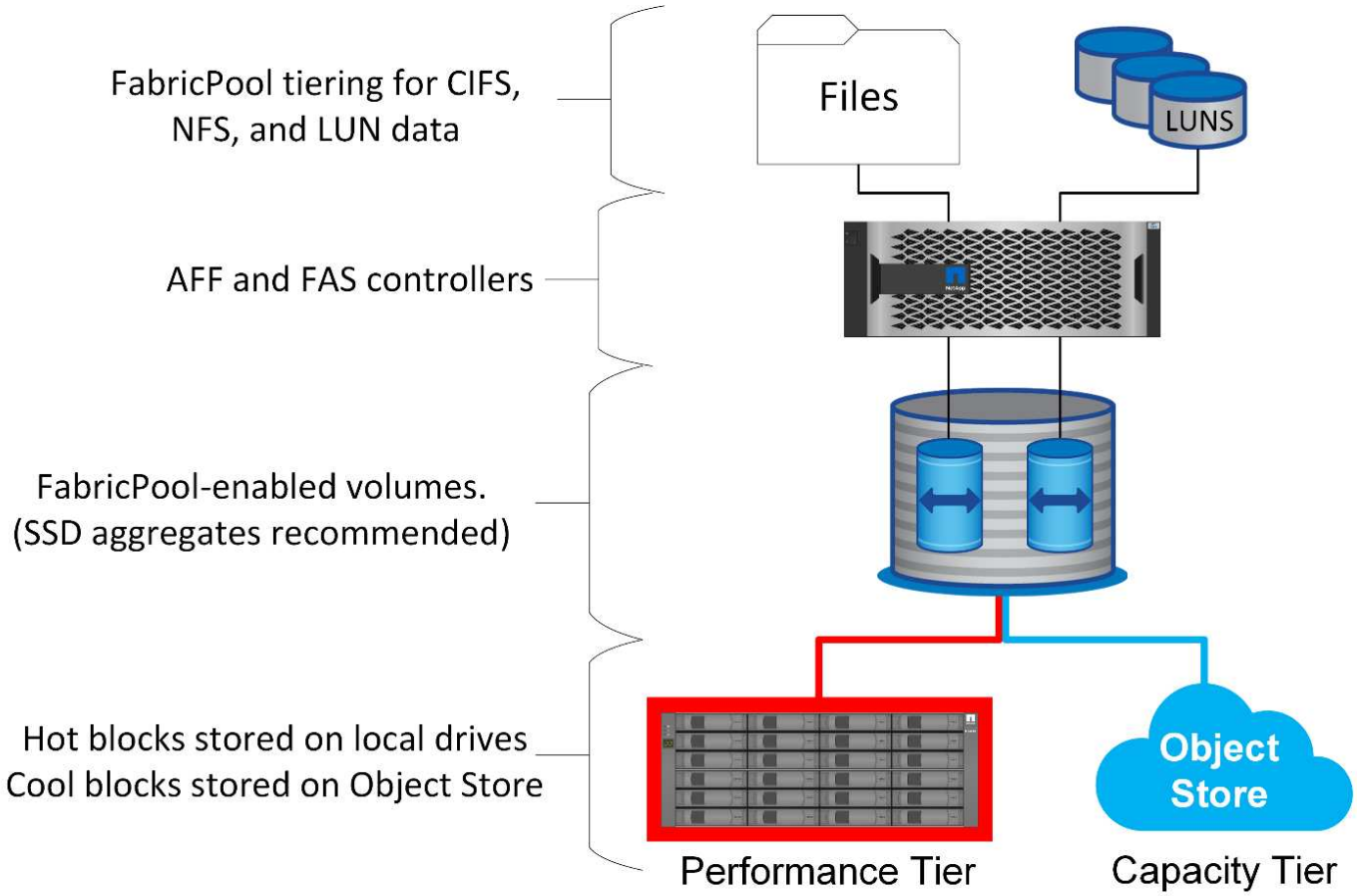
Pour comprendre l'impact du Tiering FabricPool sur Oracle et d'autres bases de données, il est nécessaire de connaître l'architecture FabricPool de bas niveau.

Architecture

FabricPool est une technologie de hiérarchisation qui classe les blocs « actifs » ou « froids » et les place dans le Tier de stockage le plus approprié. Le Tier de performance se trouve le plus souvent sur un stockage SSD et héberge les blocs de données fortement sollicités. Le Tier de capacité se trouve dans un magasin d'objets et héberge les blocs de données utiles. Elle prend en charge le stockage objet, notamment NetApp

StorageGRID, ONTAP S3, Microsoft Azure Blob Storage, le service de stockage objet Alibaba Cloud, IBM Cloud Object Storage, Google Cloud Storage et Amazon AWS S3.

Plusieurs règles de Tiering sont disponibles pour contrôler la façon dont les blocs sont classés comme actifs ou froids. Il est également possible de définir des règles par volume et de les modifier selon les besoins. Seuls les blocs de données sont déplacés entre les tiers de performance et de capacité. Les métadonnées qui définissent la structure des LUN et du système de fichiers restent toujours sur le Tier de performance. La gestion est ainsi centralisée sous ONTAP. Les fichiers et les LUN n'apparaissent pas différents des données stockées dans une autre configuration ONTAP. Le contrôleur NetApp AFF ou FAS applique les règles définies pour déplacer les données vers le Tier approprié.



Fournisseurs de magasins d'objets

Les protocoles de stockage objet utilisent de simples requêtes HTTP ou HTTPS pour stocker un grand nombre d'objets de données. L'accès au stockage objet doit être fiable, car l'accès aux données depuis ONTAP dépend du traitement rapide des demandes. Notamment Amazon S3 Standard et Infrequent Access, Microsoft Azure Hot Blob Storage, IBM Cloud et Google Cloud. Les options d'archivage telles qu'Amazon Glacier et Amazon Archive ne sont pas prises en charge, car le temps nécessaire à la récupération des données peut dépasser les tolérances des systèmes d'exploitation et des applications hôtes.

NetApp StorageGRID est également pris en charge et constitue une solution optimale. C'est un système de stockage objet haute performance, évolutif et hautement sécurisé qui assure une redondance géographique pour les données FabricPool ainsi que pour les autres applications de magasin d'objets qui font de plus en plus partie des environnements applicatifs d'entreprise.

StorageGRID peut également réduire les coûts en évitant les frais de sortie imposés par de nombreux fournisseurs de cloud public pour la lecture des données de leurs services.

Données et métadonnées

Notez que le terme « données » s'applique ici aux blocs de données réels, et non aux métadonnées. Seuls les blocs de données sont hiérarchisés, tandis que les métadonnées restent dans le Tier de performance. En outre, l'état d'un bloc en tant que bloc chaud ou froid n'est affecté que par la lecture du bloc de données réel. La simple lecture du nom, de l'horodatage ou des métadonnées de propriété d'un fichier n'affecte pas l'emplacement des blocs de données sous-jacents.

Sauvegardes

Même si FabricPool permet de réduire considérablement l'encombrement du stockage, il ne s'agit pas à lui seul d'une solution de sauvegarde. Les métadonnées NetApp WAFL restent toujours sur le Tier de performance. Si un incident catastrophique détruit le Tier de performance, il est impossible de créer un nouvel environnement à l'aide des données du Tier de capacité, car il ne contient pas de métadonnées WAFL.

FabricPool peut cependant faire partie d'une stratégie de sauvegarde. Par exemple, FabricPool peut être configuré avec la technologie de réplication NetApp SnapMirror. Chaque moitié du miroir peut avoir sa propre connexion à une cible de stockage objet. Vous obtenez ainsi deux copies indépendantes des données. La copie principale se compose des blocs du niveau de performance et des blocs associés du niveau de capacité, tandis que la réplique constitue un second ensemble de blocs de performance et de capacité.

Règles de hiérarchisation

Règles de hiérarchisation

Quatre règles sont disponibles dans ONTAP, qui contrôlent la façon dont les données Oracle du niveau de performance deviennent candidates à la relocalisation vers le niveau de capacité.

Copies Snapshot uniquement

Le `snapshot-only` `tiering-policy` s'applique uniquement aux blocs qui ne sont pas partagés avec le système de fichiers actif. Elle entraîne essentiellement une hiérarchisation des sauvegardes de bases de données. Les blocs deviennent candidats au Tiering après la création d'une copie Snapshot et l'écrasement du bloc, ce qui entraîne l'affichage d'un bloc uniquement dans la copie Snapshot. Le délai avant un `snapshot-only` le bloc est considéré comme froid est contrôlé par le `tiering-minimum-cooling-days` réglage du volume. La plage à partir de ONTAP 9.8 est de 2 à 183 jours.

De nombreux jeux de données ont des taux de modification faibles, ce qui permet de réduire au minimum les économies réalisées grâce à cette règle. Par exemple, un taux de modification hebdomadaire d'une base de données type observée sur ONTAP est inférieur à 5 %. Les journaux d'archivage de base de données peuvent occuper un espace important, mais ils continuent généralement d'exister dans le système de fichiers actif et ne sont donc pas candidats à la hiérarchisation dans le cadre de cette règle.

Auto

Le `auto` la règle de tiering étend le tiering aux blocs spécifiques de snapshot et aux blocs dans le système de fichiers actif. Le délai avant qu'un bloc soit considéré comme froid est contrôlé par le `tiering-minimum-cooling-days` réglage du volume. La plage à partir de ONTAP 9.8 est de 2 à 183 jours.

Cette approche permet d'activer des options de hiérarchisation qui ne sont pas disponibles avec le `snapshot-only` politique. Par exemple, une règle de protection des données peut nécessiter la conservation de 90 jours de certains fichiers journaux. Si vous définissez une période de refroidissement de 3 jours, tous les fichiers journaux de plus de 3 jours doivent être placés hors de la couche de performances. Cela libère un

espace considérable sur le Tier de performance tout en vous permettant de consulter et de gérer l'ensemble des 90 jours de données.

Aucune

Le `none` la règle de tiering empêche tout bloc supplémentaire d'être hiérarchisé de la couche de stockage, mais toutes les données qui se trouvent toujours dans le tier de capacité restent dans le tier de capacité jusqu'à ce qu'elles soient lues. Si le bloc est ensuite lu, il est retiré et placé sur le Tier de performance.

La principale raison d'utiliser le `none` la règle de tiering consiste à empêcher les blocs d'être hiérarchisés, mais elle peut s'avérer utile pour modifier les règles au fil du temps. Par exemple, imaginons qu'un dataset spécifique soit beaucoup hiérarchisé vers la couche de capacité, mais qu'un besoin inattendu de fonctionnalités de performance complètes se produit. La règle peut être modifiée pour éviter tout Tiering supplémentaire et confirmer que tous les blocs lus en cas d'augmentation des E/S restent dans le Tier de performance.

Tout

Le `all` la règle de tiering remplace la `backup` Politique à partir de ONTAP 9.6. Le `backup` Règle appliquée uniquement aux volumes de protection des données, c'est-à-dire une destination SnapMirror ou NetApp SnapVault. Le `all` les règles fonctionnent de même, mais ne se limitent pas aux volumes de protection des données.

Avec cette règle, les blocs sont immédiatement considérés comme « cool » et peuvent être immédiatement hiérarchisés jusqu'à la couche de capacité.

Cette règle est particulièrement appropriée pour les sauvegardes à long terme. Il peut également être utilisé comme une forme de gestion hiérarchique du stockage (HSM). Auparavant, HSM était couramment utilisé pour classer les blocs de données d'un fichier sur bande tout en gardant le fichier lui-même visible sur le système de fichiers. Un volume FabricPool avec `all` cette stratégie vous permet de stocker des fichiers dans un espace visible et gérable, tout en ne consommant quasiment aucun espace sur le niveau de stockage local.

Stratégies de récupération

Les règles de Tiering contrôlent quels blocs de base de données Oracle sont hiérarchisés du niveau de performance au niveau de capacité. Les règles de récupération contrôlent ce qui se passe lorsqu'un bloc qui a été hiérarchisé est lu.

Valeur par défaut

Tous les volumes FabricPool sont initialement définis sur `default`, ce qui signifie que le comportement est contrôlé par la `cloud-retrieval-policy. `le comportement exact dépend de la règle de hiérarchisation utilisée.

- `auto`- récupérer uniquement les données lues de façon aléatoire
- `snapshot-only`- récupérer toutes les données lues de manière séquentielle ou aléatoire
- `none`- récupérer toutes les données lues de manière séquentielle ou aléatoire
- `all`- ne pas récupérer les données du niveau de capacité

En lecture

Réglage `cloud-retrieval-policy` la lecture remplace le comportement par défaut, de sorte qu'une lecture de toutes les données hiérarchisées entraîne le renvoi de ces données vers le niveau de performance.

Par exemple, un volume peut avoir été légèrement utilisé pendant une longue période sous le `auto` la règle de tiering et la plupart des blocs sont désormais hiérarchisés.

Si une modification inattendue des besoins de l'entreprise nécessitait l'analyse répétée de certaines données pour préparer un rapport spécifique, il peut être souhaitable de modifier le `cloud-retrieval-policy` à `on-read` pour garantir que toutes les données lues sont renvoyées au niveau de performances, y compris les données lues de manière séquentielle et aléatoire. Cela améliorerait les performances des E/S séquentielles par rapport au volume.

Promouvoir

Le comportement de la règle de promotion dépend de la règle de hiérarchisation. Si la règle de hiérarchisation est `auto`, puis réglage du `cloud-retrieval-policy `to `promote` ramène tous les blocs du tier de capacité à l'analyse de tiering suivante.

Si la règle de hiérarchisation est `snapshot-only`, les seuls blocs renvoyés sont les blocs associés au système de fichiers actif. Normalement, cela n'aurait aucun effet car les seuls blocs placés sous le sont `snapshot-only` la règle serait les blocs associés exclusivement aux snapshots. Il n'y aurait pas de blocs hiérarchisés dans le système de fichiers actif.

Toutefois, si une SnapRestore de volume ou une opération de clonage de fichiers a été effectuée pour restaurer les données d'un volume à partir d'un snapshot, le système de fichiers actif peut désormais avoir besoin de certains blocs qui ont été hiérarchisés, car ils n'étaient associés qu'à des snapshots. Il peut être souhaitable de modifier temporairement le `cloud-retrieval-policy` règle à `promote` pour récupérer rapidement tous les blocs localement requis.

Jamais

Ne récupérez pas les blocs du niveau de capacité.

Stratégies de Tiering

Tiering complet de fichiers

Bien que le Tiering FabricPool fonctionne au niveau des blocs, il peut dans certains cas servir à fournir un Tiering au niveau des fichiers.

De nombreux jeux de données d'applications sont organisés par date, et ces données sont généralement moins susceptibles d'être accessibles au fur et à mesure du vieillissement. Par exemple, une banque peut disposer d'un référentiel de fichiers PDF contenant cinq années de relevés clients, mais seuls les derniers mois sont actifs. FabricPool peut être utilisé pour déplacer d'anciens fichiers de données vers le Tier de capacité. Une période de refroidissement de 14 jours permettrait de conserver les fichiers PDF de 14 jours les plus récents sur le niveau de performance. En outre, les fichiers lus au moins tous les 14 jours resteraient fortement sollicités et resteraient donc sur le Tier de performance.

Stratégies

Pour mettre en œuvre une approche de hiérarchisation basée sur des fichiers, vous devez avoir des fichiers écrits et non modifiés par la suite. Le `tiering-minimum-cooling-days` la règle doit être définie suffisamment haut pour que les fichiers dont vous avez besoin restent sur le tier de performance. Par exemple, un jeu de données pour lequel les 60 derniers jours de données sont requis avec des performances optimales garantit le paramétrage du `tiering-minimum-cooling-days` période jusqu'en 60. Des résultats similaires peuvent également être obtenus en fonction des modèles d'accès aux fichiers. Par exemple, si les 90 derniers jours de données sont requis et que l'application accède à cette période de 90 jours, les données restent sur le

Tier de performance. En réglant le `tiering-minimum-cooling-days` sur 2, le tiering s'affiche rapidement une fois les données moins actives.

Le `auto` la règle est requise pour la hiérarchisation de ces blocs, car uniquement le système `auto` la règle affecte les blocs qui se trouvent dans le système de fichiers actif.

> Tout type d'accès aux données réinitialise les données de la carte thermique. L'analyse antivirus, l'indexation et même l'activité de sauvegarde qui lit les fichiers source empêchent le Tiering, car les besoins sont importants `tiering-minimum-cooling-days` le seuil n'est jamais atteint.

Tiering partiel des fichiers

Comme FabricPool fonctionne au niveau des blocs, les fichiers susceptibles d'être modifiés peuvent être partiellement hiérarchisés vers un stockage objet tout en restant partiellement sur le Tier de performance.

Ceci est courant avec les bases de données. Les bases de données qui contiennent des blocs inactifs sont également candidates au Tiering FabricPool. Par exemple, une base de données de gestion de la chaîne logistique peut contenir des informations historiques qui doivent être disponibles si nécessaire, mais qui ne sont pas accessibles pendant les opérations normales. FabricPool peut être utilisé pour déplacer de manière sélective les blocs inactifs.

Par exemple, les fichiers de données s'exécutant sur un volume FabricPool avec un `tiering-minimum-cooling-days` la période de 90 jours permet de conserver les blocs auxquels le tier de performance accède au cours des 90 jours précédents. Toutefois, tout élément non utilisé pendant 90 jours est transféré vers le niveau de capacité. Dans d'autres cas, l'activité normale de l'application préserve les blocs corrects du niveau approprié. Par exemple, si une base de données est normalement utilisée pour traiter les 60 jours précédents de données sur une base régulière, c'est beaucoup moins `tiering-minimum-cooling-days` la période peut être définie car l'activité naturelle de l'application s'assure que les blocs ne sont pas déplacés prématurément.

> Le `auto` la politique doit être utilisée avec soin pour les bases de données. De nombreuses bases de données ont des activités périodiques, comme le processus de fin de trimestre ou les opérations de réindexation. Si la période de ces opérations est supérieure à `tiering-minimum-cooling-days` des problèmes de performances peuvent se produire. Par exemple, si le traitement de fin de trimestre nécessite 1 To de données qui n'étaient pas modifiées, ces données peuvent maintenant être présentes sur le niveau de capacité. Les lectures à partir du niveau de capacité sont souvent extrêmement rapides et ne provoquent pas de problèmes de performance, mais les résultats exacts dépendent de la configuration du magasin d'objets.

Stratégies

Le `tiering-minimum-cooling-days` la règle doit être définie de manière suffisamment élevée pour conserver les fichiers qui peuvent être requis sur le niveau de performance. Par exemple, une base de données dans laquelle les 60 derniers jours de données peuvent être requis avec des performances optimales justifierait de définir le `tiering-minimum-cooling-days` période à 60 jours. Des résultats similaires pourraient également être obtenus en fonction des modèles d'accès aux fichiers. Par exemple, si les 90 derniers jours de données sont requis et que l'application accède à cette période de 90 jours, les données restent sur le Tier de performance. Réglage du `tiering-minimum-cooling-days` une période de 2 jours permettrait de hiérarchiser les données rapidement lorsque celles-ci deviennent moins actives.

Le `auto` la règle est requise pour la hiérarchisation de ces blocs, car uniquement le système `auto` la règle affecte les blocs qui se trouvent dans le système de fichiers actif.

> Tout type d'accès aux données réinitialise les données de la carte thermique. Par conséquent, les analyses de la table complète des bases de données, et même les opérations de sauvegarde qui lisent les fichiers source, empêchent le Tiering, car nécessaire `tiering-minimum-cooling-days` le seuil n'est jamais atteint.

Tiering des journaux d'archivage

L'utilisation la plus importante pour FabricPool est peut-être l'amélioration de l'efficacité des données inactives connues, telles que les journaux de transactions de base de données.

La plupart des bases de données relationnelles opèrent en mode d'archivage du journal de transactions pour assurer une restauration instantanée. Les modifications apportées aux bases de données sont validées en enregistrant les modifications dans les journaux de transactions et le journal de transactions est conservé sans être écrasé. Il peut donc s'avérer nécessaire de conserver un énorme volume de journaux de transactions archivés. De nombreux autres workflows applicatifs génèrent des données qui doivent être conservées, mais il est très peu probable qu'elles soient accessibles.

Pour résoudre ces problèmes, FabricPool propose une solution unique avec hiérarchisation intégrée. Les fichiers sont stockés et restent accessibles à leur emplacement habituel, mais ne prennent pratiquement pas d'espace sur la baie principale.

Stratégies

Utiliser un `tiering-minimum-cooling-days` la règle de quelques jours permet de conserver les blocs dans les fichiers récemment créés (les fichiers les plus susceptibles d'être requis à court terme) sur le niveau de performance. Les blocs de données des anciens fichiers sont ensuite déplacés vers le niveau de capacité.

Le `auto` applique la hiérarchisation des invites lorsque le seuil de refroidissement a été atteint, que les journaux aient été supprimés ou qu'ils continuent d'exister dans le système de fichiers principal. Le stockage de tous les journaux potentiellement requis dans un seul emplacement du système de fichiers actif simplifie également la gestion. Il n'y a aucune raison de rechercher un fichier à restaurer à l'aide de snapshots.

Certaines applications, telles que Microsoft SQL Server, tronquent les fichiers journaux de transactions pendant les opérations de sauvegarde afin que les journaux ne soient plus dans le système de fichiers actif. Il est possible d'économiser de la capacité à l'aide de `snapshot-only` la règle de tiering, mais la règle `auto` la règle n'est pas utile pour les données de journal car il devrait rarement y avoir des données de journal refroidies dans le système de fichiers actif.

Tiering Snapshot

La version initiale de FabricPool a ciblé le cas d'utilisation de la sauvegarde. Les seuls types de blocs qui ont pu être hiérarchisés sont les blocs qui n'étaient plus associés aux données dans le système de fichiers actif. Par conséquent, seuls les blocs de données des snapshots peuvent être déplacés vers le niveau de capacité. Il s'agit là de l'une des options de hiérarchisation les plus sécurisées lorsque vous devez vous assurer que les performances ne sont jamais affectées.

Règles - snapshots locaux

Deux options sont disponibles pour le Tiering des blocs de snapshots inactifs vers le niveau de capacité. Tout d'abord, le `snapshot-only` la règle cible uniquement les blocs de snapshot. Bien que le `auto` la politique inclut le `snapshot-only` et tiering des blocs à partir du système de fichiers actif. Ce n'est peut-être pas souhaitable.

Le `tiering-minimum-cooling-days` value doit être défini sur une période qui met à disposition les données éventuellement requises lors d'une restauration sur le tier de performance. Par exemple, la plupart des scénarios de restauration d'une base de données de production stratégique incluent un point de restauration à un moment donné au cours des jours précédents. Réglage a `tiering-minimum-cooling-days` la valeur 3 garantit que toute restauration du fichier entraîne un fichier qui offre immédiatement des performances maximales. Tous les blocs des fichiers actifs sont toujours présents sur un système de stockage rapide sans avoir à les restaurer à partir du niveau de capacité.

Règles - snapshots répliqués

Les snapshots répliqués avec SnapMirror ou SnapVault, uniquement utilisés pour la restauration, doivent généralement utiliser FabricPool `all` politique. Avec cette règle, les métadonnées sont répliquées, mais tous les blocs de données sont immédiatement envoyés au niveau de capacité pour des performances maximales. La plupart des processus de restauration impliquent des E/S séquentielles, ce qui est intrinsèquement efficace. Le délai de restauration à partir de la destination du magasin d'objets doit être évalué, mais dans une architecture bien conçue, ce processus de restauration ne doit pas nécessairement être beaucoup plus lent que la restauration à partir de données locales.

Si les données répliquées sont également destinées à être utilisées pour le clonage, le `auto` la politique est plus appropriée, avec un `tiering-minimum-cooling-days` valeur qui englobe les données qui doivent être utilisées régulièrement dans un environnement de clonage. Par exemple, le jeu de travail actif d'une base de données peut inclure des données lues ou écrites au cours des trois jours précédents, mais il peut également inclure 6 mois de données historiques supplémentaires. Si oui, alors le `auto` La règle appliquée à la destination SnapMirror met à disposition le jeu de travail sur le Tier de performance.

Tiering de sauvegarde

Les sauvegardes d'applications traditionnelles incluent des produits tels qu'Oracle Recovery Manager, qui créent des sauvegardes basées sur des fichiers en dehors de l'emplacement de la base de données d'origine.

```
 `tiering-minimum-cooling-days` policy of a few days preserves the most
recent backups, and therefore the backups most likely to be required for
an urgent recovery situation, on the performance tier. The data blocks of
the older files are then moved to the capacity tier.
Le `auto` la règle est la règle la plus appropriée pour les données de
sauvegarde. Cela garantit une hiérarchisation rapide lorsque le seuil de
refroidissement a été atteint, que les fichiers aient été supprimés ou
qu'ils continuent d'exister dans le système de fichiers principal. Le
stockage de tous les fichiers potentiellement requis dans un emplacement
unique du système de fichiers actif simplifie également la gestion. Il n'y
a aucune raison de rechercher un fichier à restaurer à l'aide de
snapshots.
```

Le `snapshot-only` la stratégie peut être mise en œuvre, mais elle s'applique uniquement aux blocs qui ne sont plus dans le système de fichiers actif. Par conséquent, les fichiers d'un partage NFS ou SMB doivent d'abord être supprimés avant de pouvoir placer les données dans un Tier.

Cette règle serait encore moins efficace avec une configuration de LUN, car la suppression d'un fichier d'une LUN supprime uniquement les références de fichier des métadonnées du système de fichiers. Les blocs réels des LUN restent en place jusqu'à ce qu'ils soient remplacés. Cette situation peut entraîner un délai long entre la suppression d'un fichier et l'écrasement des blocs et leur candidature à la hiérarchisation. Il est avantageux de déplacer le `snapshot-only` Bloque le niveau de capacité, mais, dans l'ensemble, la gestion FabricPool des données de sauvegarde fonctionne mieux avec le `auto` politique.

> Cette approche permet aux utilisateurs de gérer plus efficacement l'espace requis pour les sauvegardes, mais FabricPool lui-même n'est pas une technologie de sauvegarde. Le Tiering des fichiers de sauvegarde vers un magasin d'objets simplifie la gestion, car les fichiers restent visibles sur le système de stockage d'origine. Cependant, les blocs de données de destination du magasin d'objets dépendent du système de stockage d'origine. En cas de perte du volume source, les données du magasin d'objets ne sont plus utilisables.

Interruptions d'accès au magasin d'objets

Le Tiering d'un dataset avec FabricPool entraîne une dépendance entre la baie de stockage primaire et le Tier de magasin d'objets. De nombreuses options de stockage objet offrent différents niveaux de disponibilité. Il est important de comprendre l'impact d'une éventuelle perte de connectivité entre la baie de stockage primaire et le niveau de stockage objet.

Si une E/S émise par ONTAP nécessite des données du niveau de capacité et que les ONTAP ne peuvent pas atteindre le niveau de capacité pour récupérer des blocs, les E/S finissent par être sorties. L'effet de ce délai dépend du protocole utilisé. Dans un environnement NFS, ONTAP répond par une réponse EJUKEBOX ou EDELAY, selon le protocole. Certains systèmes d'exploitation plus anciens peuvent interpréter cela comme une erreur, mais les systèmes d'exploitation actuels et les niveaux de correctifs actuels du client Oracle Direct NFS traitent cette erreur comme une nouvelle tentative et continuent d'attendre la fin des E/S.

Un délai plus court s'applique aux environnements SAN. Si un bloc de l'environnement de magasin d'objets est requis et reste inaccessible pendant deux minutes, une erreur de lecture est renvoyée à l'hôte. Le volume ONTAP et les LUN restent en ligne, mais le système d'exploitation hôte peut signaler le système de fichiers comme étant dans un état d'erreur.

Les problèmes de connectivité du stockage objet `snapshot-only` la politique est moins préoccupante, car seules les données de sauvegarde sont hiérarchisées. Les problèmes de communication ralentiraient la récupération des données, mais n'affecteraient pas les données utilisées activement. Le `auto` et `all` Les règles permettent le Tiering des données inactives de la LUN active, ce qui signifie qu'une erreur lors de la récupération des données du magasin d'objets peut affecter la disponibilité de la base de données. Un déploiement SAN doté de ces règles doit uniquement être utilisé avec un stockage objet de grande qualité et des connexions réseau conçues pour une haute disponibilité. NetApp StorageGRID est la meilleure option.

Protection des données Oracle

Protection des données avec ONTAP

NetApp sait que les bases de données contient les données les plus stratégiques.

Une entreprise ne peut pas fonctionner sans accéder à ses données, et parfois l'activité repose sur les données. Ces données doivent être protégées, mais la protection ne se limite pas à garantir une sauvegarde utilisable. Elle consiste également à effectuer des sauvegardes rapidement et de manière fiable en plus de les stocker en toute sécurité.

L'autre côté de la protection des données est la restauration des données. Lorsque les données ne sont pas accessibles, l'entreprise est affectée et peut ne pas fonctionner tant qu'elle n'est pas restaurée. Ce processus doit être rapide et fiable. Enfin, la plupart des bases de données doivent être protégées contre les incidents, ce qui signifie maintenir une réplique de la base de données. La réplique doit être suffisamment à jour. Il doit également être rapide et simple de faire de la réplique une base de données entièrement opérationnelle.

> Cette documentation remplace le rapport technique *TR-4591 : protection des données Oracle : sauvegarde, restauration et réplication.*

Planification

Une architecture de protection des données d'entreprise adaptée dépend des exigences de l'entreprise concernant la conservation des données, la restauration et la tolérance aux perturbations à divers moments.

Prenons l'exemple du nombre d'applications, de bases de données et de datasets importants pris en compte. Il est relativement simple d'élaborer une stratégie de sauvegarde pour un seul dataset afin d'assurer la conformité aux SLA standard, car la gestion ne comporte pas beaucoup d'objets. À mesure que le nombre de jeux de données augmente, la surveillance devient plus complexe et les administrateurs peuvent être obligés de consacrer de plus en plus de temps aux pannes de sauvegarde. Dès qu'un environnement évolue, il faut adopter une approche totalement différente.

La taille des datasets affecte également la stratégie. Par exemple, le jeu de données étant si petit, de nombreuses options sont possibles pour la sauvegarde et la restauration avec une base de données de 100 Go. En général, la simple copie des données à partir du support de sauvegarde avec des outils classiques permet d'atteindre un RTO suffisant pour la restauration. Une base de données de 100 To a généralement besoin d'une stratégie totalement différente, sauf si le RTO autorise une panne de plusieurs jours. Dans ce cas, une procédure classique de sauvegarde et de restauration basée sur des copies peut être acceptable.

Enfin, il y a des facteurs en dehors du processus de sauvegarde et de restauration lui-même. Par exemple, existe-t-il des bases de données qui prennent en charge les activités de production stratégiques, faisant de la restauration un événement rare uniquement effectué par des administrateurs de bases de données qualifiés ? Ou bien, les bases de données font-elles partie d'un vaste environnement de développement dans lequel la restauration est fréquente et gérée par une équipe INFORMATIQUE généraliste ?

Planification des objectifs de durée de restauration, de point de récupération et des SLA

ONTAP vous permet d'adapter facilement une stratégie de protection des données des bases de données Oracle aux besoins de votre entreprise.

Ces exigences comprennent des facteurs tels que la vitesse de restauration, la perte de données maximale autorisée et les besoins de conservation des sauvegardes. Le plan de protection des données doit également tenir compte de diverses exigences réglementaires en matière de conservation et de restauration des données. Enfin, différents scénarios de restauration des données doivent être pris en compte, allant de la restauration classique et prévisible résultant d'erreurs d'utilisateurs ou d'applications à des scénarios de reprise sur incident incluant la perte complète d'un site.

Les modifications mineures apportées aux règles de protection et de restauration des données peuvent avoir

un impact significatif sur l'architecture globale du stockage, de la sauvegarde et de la restauration. Il est essentiel de définir et de documenter des normes avant de commencer le travail de conception afin d'éviter de compliquer une architecture de protection des données. Des fonctions ou des niveaux de protection inutiles entraînent des coûts et des frais de gestion inutiles. Par ailleurs, une exigence initialement négligée peut conduire un projet dans la mauvaise direction ou nécessiter des modifications de conception de dernière minute.

Objectif de délai de restauration

L'objectif de délai de restauration (RTO) définit le temps maximal autorisé pour la restauration d'un service. Par exemple, une base de données de ressources humaines peut atteindre un objectif de délai de restauration de 24 heures. En effet, même s'il ne serait pas très pratique de perdre l'accès à ces données pendant les jours de travail, l'entreprise peut tout de même fonctionner. En revanche, une base de données prenant en charge le grand livre d'une banque aurait un RTO mesuré en minutes, voire en secondes. Un objectif RTO de zéro n'est pas possible, car il doit y avoir un moyen de faire la différence entre une panne de service réelle et un événement de routine tel qu'un paquet réseau perdu. Toutefois, un objectif RTO quasi nul est généralement requis.

Objectif de point de récupération

L'objectif de point de récupération (RPO) définit la perte de données maximale tolérable. Dans de nombreux cas, l'objectif de point de récupération est uniquement déterminé par la fréquence des copies Snapshot ou des mises à jour snapmirror.

Dans certains cas, le RPO peut être rendu plus agressif, car il permet de protéger certaines données de manière sélective plus fréquemment. Dans un contexte de base de données, le RPO correspond généralement à la quantité de données perdues dans un journal spécifique. Dans un scénario de restauration typique dans lequel une base de données est endommagée en raison d'un bogue de produit ou d'une erreur utilisateur, le RPO doit être égal à zéro, ce qui signifie qu'il ne doit pas y avoir de perte de données. La procédure de restauration implique la restauration d'une copie antérieure des fichiers de base de données, puis la relecture des fichiers journaux pour ramener l'état de la base de données au point dans le temps souhaité. Les fichiers journaux requis pour cette opération doivent déjà être en place à l'emplacement d'origine.

Dans des scénarios inhabituels, les données des journaux peuvent être perdues. Par exemple, un accident ou un acte malveillant `rm -rf *` des fichiers de base de données peuvent entraîner la suppression de toutes les données. La seule option serait de restaurer des données à partir de sauvegardes, y compris des fichiers journaux, et certaines seraient inévitablement perdues. Dans un environnement de sauvegarde classique, la seule option permettant d'améliorer le RPO consiste à effectuer des sauvegardes répétées des données du journal. Cela a toutefois ses limites en raison du déplacement constant des données et de la difficulté à maintenir un système de sauvegarde en tant que service en continu. L'un des avantages des systèmes de stockage avancés est la capacité à protéger les données contre les dommages accidentels ou malveillants aux fichiers et à fournir ainsi un meilleur RPO sans déplacement des données.

Reprise après incident

La reprise après incident comprend l'architecture INFORMATIQUE, les règles et les procédures requises pour restaurer un service en cas d'incident physique. Cela peut inclure les inondations, les incendies ou les personnes agissant avec une intention malveillante ou négligente.

La reprise sur incident est bien plus qu'un ensemble de procédures de restauration. Il s'agit du processus complet d'identification des différents risques, de définition des exigences en matière de restauration des données et de continuité des services, et de mise à disposition de l'architecture appropriée avec les procédures associées.

Lors de l'établissement des exigences de protection des données, il est essentiel de faire la différence entre les objectifs RPO et RTO types et les exigences RPO et RTO requises pour la reprise après incident. Pour les situations de perte de données, allant d'une erreur utilisateur relativement normale à un incendie qui détruit un data Center, certains environnements applicatifs nécessitent un RPO nul et un RTO quasi nul. Cependant, il y a des conséquences administratives et des coûts pour ces niveaux élevés de protection.

En général, les exigences de restauration des données non liées aux incidents doivent être strictes pour deux raisons. Tout d'abord, les bogues d'application et les erreurs d'utilisateur qui endommagent les données sont prévisibles au point qu'ils sont presque inévitables. Deuxièmement, il n'est pas difficile de concevoir une stratégie de sauvegarde capable de fournir un RPO nul et un RTO faible tant que le système de stockage n'est pas détruit. Il n'y a aucune raison de ne pas traiter un risque important facilement résolu. C'est pourquoi les objectifs RPO et RTO pour la reprise locale doivent être agressifs.

Les exigences en termes de RTO et de RPO pour la reprise d'activité varient plus largement en fonction du risque d'incident et des conséquences de la perte de données ou de l'interruption pour une entreprise. Les exigences en matière de RPO et de RTO doivent être basées sur les besoins réels de l'entreprise et non sur des principes généraux. Ils doivent prendre en compte plusieurs scénarios de catastrophe physique et logique.

Incidents logiques

Les incidents logiques incluent la corruption des données provoquée par les utilisateurs, les bogues des applications ou du système d'exploitation et les dysfonctionnements logiciels. Les incidents logiques peuvent également inclure des attaques malveillantes de tiers contenant des virus ou des vers, ou encore en exploitant les vulnérabilités des applications. Dans ces cas, l'infrastructure physique n'est pas endommagée, mais les données sous-jacentes ne sont plus valides.

Les ransomwares sont un type de catastrophe logique de plus en plus courant qui sert à chiffrer les données à l'aide d'un vecteur d'attaque. Le chiffrement n'endommage pas les données, mais il les rend indisponibles jusqu'à ce que le paiement soit effectué à un tiers. De plus en plus d'entreprises sont spécifiquement la cible de piratage. Face à cette menace, NetApp propose des snapshots inviolables où même l'administrateur du stockage ne peut pas modifier les données protégées avant la date d'expiration configurée.

Incidents physiques

Les incidents physiques incluent la défaillance de composants d'une infrastructure qui dépasse ses capacités de redondance et entraînent une perte de données ou une perte de service prolongée. Par exemple, la protection RAID assure la redondance des disques durs et l'utilisation de HBA assure la redondance des ports FC et des câbles FC. Les pannes matérielles de ces composants sont prévisibles et n'ont pas d'incidence sur la disponibilité.

Dans un environnement d'entreprise, il est généralement possible de protéger l'infrastructure d'un site entier avec des composants redondants au point où le seul scénario de catastrophe physique prévisible est la perte complète du site. La planification de la reprise d'activité dépend alors de la réplication de site à site.

Protection des données synchrone et asynchrone

Dans l'idéal, toutes les données seraient répliquées de manière synchrone sur des sites dispersés géographiquement. Une telle réplication n'est pas toujours possible, voire possible pour plusieurs raisons :

- La réplication synchrone entraîne inévitablement une augmentation de la latence d'écriture, car toutes les modifications doivent être répliquées vers les deux emplacements avant que l'application/la base de données ne puisse poursuivre le traitement. L'effet de performance qui en résulte est parfois inacceptable, excluant l'utilisation de la mise en miroir synchrone.
- En raison de l'adoption accrue de 100 % de stockage SSD, il est plus probable que l'on remarque une latence d'écriture supplémentaire, car les attentes en termes de performances comprennent des centaines

de milliers d'IOPS et une latence inférieure à la milliseconde. Pour tirer pleinement parti de l'utilisation de 100 % des SSD, il peut être nécessaire de revoir la stratégie de reprise sur incident.

- La croissance des datasets en octets continue, ce qui engendre des défis en garantissant une bande passante suffisante pour soutenir la réplication synchrone.
- La croissance des datasets s'accompagne également de défis liés à la gestion de la réplication synchrone à grande échelle.
- Les stratégies basées sur le cloud impliquent souvent des distances de réplication et une latence plus importantes, ce qui exclut davantage l'utilisation de la mise en miroir synchrone.

NetApp propose des solutions qui incluent à la fois la réplication synchrone pour satisfaire les besoins les plus exigeants en matière de restauration des données et des solutions asynchrones qui assurent des performances et une flexibilité accrues. De plus, la technologie NetApp s'intègre en toute transparence à de nombreuses solutions de réplication tierces, telles qu'Oracle DataGuard

Durée de conservation

Le dernier aspect d'une stratégie de protection des données est la durée de conservation des données, qui peut varier considérablement.

- Il est généralement nécessaire d'effectuer 14 jours de sauvegardes nocturnes sur le site principal et 90 jours de sauvegardes sur un site secondaire.
- De nombreux clients créent des archives trimestrielles autonomes stockées sur différents supports.
- Une base de données constamment mise à jour n'a peut-être pas besoin de données historiques, et les sauvegardes ne doivent être conservées que pendant quelques jours.
- Pour des raisons réglementaires, une capacité de restauration peut être nécessaire au point de toute transaction arbitraire dans une fenêtre de 365 jours.

Disponibilité de la base de données

ONTAP est conçu pour offrir une disponibilité maximale des bases de données Oracle. Ce document ne contient pas de description complète des fonctionnalités de haute disponibilité de ONTAP. Cependant, comme pour la protection des données, il est important de bien comprendre cette fonctionnalité lors de la conception d'une infrastructure de base de données.

Paires HA

L'unité de base de la haute disponibilité est la paire haute disponibilité. Chaque paire contient des liens redondants pour prendre en charge la réplication des données vers la mémoire NVRAM. La NVRAM n'est pas un cache d'écriture. La RAM à l'intérieur du contrôleur sert de cache d'écriture. L'objectif de la mémoire NVRAM est de journaliser temporairement les données afin de prévenir toute panne système inattendue. À cet égard, il est similaire à un fichier redo log de base de données.

La mémoire NVRAM et le journal de reprise de base de données sont utilisés pour stocker des données rapidement, ce qui permet d'y apporter les modifications le plus rapidement possible. La mise à jour des données persistantes sur les disques (ou fichiers de données) n'a lieu qu'une fois plus tard lors d'un processus appelé point de contrôle sur ONTAP et la plupart des plateformes de bases de données. Les données NVRAM et les redo logs de base de données ne sont pas lus pendant les opérations normales.

Si un contrôleur tombe en panne brusquement, des modifications sont susceptibles d'être en attente de stockage dans la mémoire NVRAM qui n'ont pas encore été écrites sur les disques. Le contrôleur partenaire

détecte la panne, prend le contrôle des disques et applique les modifications requises qui ont été stockées dans la mémoire NVRAM.

Takeover et Giveback

Le basculement et le rétablissement font référence au processus de transfert de la responsabilité des ressources de stockage entre les nœuds d'une paire HA. Le basculement et le rétablissement sont deux aspects :

- Gestion de la connectivité réseau permettant l'accès aux lecteurs
- Gestion des disques eux-mêmes

Les interfaces réseau prenant en charge le trafic CIFS et NFS sont configurées avec un emplacement de home et de basculement. Il inclut le déplacement des interfaces réseau vers leur domicile temporaire sur une interface physique située sur le(s) même(s) sous-réseau que l'emplacement d'origine. Le rétablissement inclut le déplacement des interfaces réseau vers leurs emplacements d'origine. Le comportement exact peut être réglé selon les besoins.

Les interfaces réseau prenant en charge les protocoles de bloc SAN, tels que iSCSI et FC, ne sont pas déplacées pendant le basculement et le rétablissement. Les LUN doivent plutôt être provisionnées avec des chemins qui incluent une paire HA complète entraînant un chemin principal et un chemin secondaire.

> Des chemins d'accès supplémentaires vers des contrôleurs supplémentaires peuvent également être configurés pour prendre en charge le déplacement des données entre les nœuds d'un cluster plus grand, mais cela ne fait pas partie du processus de haute disponibilité.

Le deuxième aspect du Takeover et Giveback est le transfert de la propriété de disque. Le processus exact dépend de plusieurs facteurs, notamment la raison du Takeover/Giveback et les options de ligne de commande émises. L'objectif est de réaliser l'opération aussi efficacement que possible. Bien que le processus global puisse sembler durer plusieurs minutes, le moment réel où la propriété du disque est transférée d'un nœud à un autre peut généralement se mesurer en secondes.

Temps de reprise

Les E/S de l'hôte font l'objet d'une courte pause au niveau des E/S lors des opérations de basculement et de rétablissement. Cependant, la configuration de l'environnement ne doit pas provoquer d'interruption des applications. Le processus de transition réel dans lequel les E/S sont retardées se mesure généralement en secondes, mais l'hôte peut avoir besoin de plus de temps pour reconnaître la modification des chemins de données et renvoyer les opérations d'E/S.

La nature de la perturbation dépend du protocole :

- Une interface réseau prenant en charge le trafic NFS et CIFS émet une requête ARP (Address Resolution Protocol) vers le réseau après la transition vers un nouvel emplacement physique. Les commutateurs réseau mettent ainsi à jour leurs tables d'adresses MAC (Media Access Control) et reprennent le traitement des E/S. L'interruption dans le cas d'un basculement et d'un rétablissement planifiés se mesure généralement en secondes et, dans la plupart des cas, elle n'est pas détectable. Certains réseaux peuvent être plus lents à reconnaître pleinement le changement de chemin réseau et certains systèmes d'exploitation peuvent mettre en file d'attente beaucoup d'E/S dans un délai très court qui doit être réessayé. Cela peut prolonger le temps nécessaire pour reprendre les E/S.
- Une interface réseau prenant en charge les protocoles SAN ne peut pas être mise à niveau vers un nouvel emplacement. Un système d'exploitation hôte doit modifier le ou les chemins utilisés. La pause des E/S observée par l'hôte dépend de plusieurs facteurs. Du point de vue du système de stockage, la période

pendant laquelle les E/S ne peuvent pas être servies ne prend que quelques secondes. Cependant, des systèmes d'exploitation hôtes différents peuvent nécessiter plus de temps pour permettre à une E/S de se déconnecter avant de réessayer. Les systèmes d'exploitation les plus récents sont mieux à même de reconnaître un changement de chemin beaucoup plus rapidement, mais les systèmes d'exploitation plus anciens nécessitent généralement jusqu'à 30 secondes pour reconnaître un changement.

Les délais de basculement attendus lors desquels le système de stockage ne peut pas transmettre de données à un environnement applicatif sont indiqués dans le tableau ci-dessous. Aucun environnement applicatif ne doit contenir d'erreurs ; le basculement doit alors apparaître sous forme de courte pause dans le traitement des E/S.

	NFS	AFF	ASA
Basculement planifié	15 s	6-10 s	2-3 s
Basculement non planifié	30 s	6-10 s	2-3 s

Checksums et intégrité des données

ONTAP et les protocoles qu'il prend en charge incluent de nombreuses fonctionnalités qui protègent l'intégrité des bases de données Oracle, notamment les données au repos et les données transmises sur le réseau.

La protection logique des données dans ONTAP comprend trois exigences clés :

- Les données doivent être protégées contre la corruption.
- Les données doivent être protégées contre les pannes disques.
- Les modifications de données doivent être protégées contre la perte.

Ces trois besoins sont abordés dans les sections suivantes.

Corruption du réseau : checksums

Le niveau de protection de données le plus élémentaire est la somme de contrôle, qui est un code spécial de détection d'erreur stocké avec les données. La corruption des données lors de la transmission du réseau est détectée grâce à l'utilisation d'un checksum et, dans certains cas, de multiples checksums.

Par exemple, une trame FC inclut une forme de somme de contrôle appelée contrôle de redondance cyclique (CRC) pour s'assurer que la charge utile n'est pas corrompue en transit. L'émetteur envoie les données et le CRC des données. Le récepteur d'une trame FC recalcule le CRC des données reçues pour s'assurer qu'il correspond au CRC transmis. Si le nouveau CRC calculé ne correspond pas au CRC joint à la trame, les données sont corrompues et la trame FC est supprimée ou rejetée. Une opération d'E/S iSCSI comprend des checksums au niveau des couches TCP/IP et Ethernet. Pour une protection supplémentaire, elle peut également inclure la protection CRC facultative au niveau de la couche SCSI. Toute corruption de bit sur le fil est détectée par la couche TCP ou la couche IP, ce qui entraîne la retransmission du paquet. Comme avec FC, les erreurs dans le CRC SCSI entraînent une suppression ou un rejet de l'opération.

Corruption de disque : checksums

Des checksums sont également utilisés pour vérifier l'intégrité des données stockées sur les disques. Les blocs de données écrits sur les disques sont stockés avec une fonction de checksum qui génère un nombre imprévisible lié aux données d'origine. Lorsque les données sont lues à partir du lecteur, la somme de contrôle est recalculée et comparée à la somme de contrôle stockée. Si elle ne correspond pas, les données sont

corrompues et doivent être restaurées par la couche RAID.

Corruption des données : écritures perdues

L'un des types de corruption les plus difficiles à détecter est une écriture perdue ou mal placée. Lorsqu'une écriture est reconnue, elle doit être écrite sur le support à l'emplacement correct. La corruption des données sur place est relativement facile à détecter à l'aide d'une simple somme de contrôle stockée avec les données. Cependant, si l'écriture est simplement perdue, alors la version précédente des données peut toujours exister et le total de contrôle serait correct. Si l'écriture est placée au mauvais emplacement physique, la somme de contrôle associée sera à nouveau valide pour les données stockées, même si l'écriture a détruit d'autres données.

La solution à ce défi est la suivante :

- Une opération d'écriture doit inclure des métadonnées indiquant l'emplacement où l'écriture est attendue.
- Une opération d'écriture doit inclure une sorte d'identifiant de version.

Lorsque ONTAP écrit un bloc, il inclut les données à l'emplacement où ce bloc appartient. Si une lecture ultérieure identifie un bloc, mais que les métadonnées indiquent qu'il appartient à l'emplacement 123 lorsqu'il a été trouvé à l'emplacement 456, l'écriture a été déplacée.

Il est plus difficile de détecter une écriture entièrement perdue. L'explication est très complexe, mais ONTAP stocke les métadonnées de façon à ce qu'une opération d'écriture entraîne des mises à jour vers deux emplacements différents sur les disques. En cas de perte d'une écriture, une lecture ultérieure des données et des métadonnées associées affiche deux identités de version différentes. Cela indique que l'écriture n'a pas été effectuée par le lecteur.

La corruption des écritures perdues ou déplacées est extrêmement rare. Cependant, avec la croissance continue des disques et l'expansion des jeux de données en exaoctets, le risque augmente. La détection des pertes en écriture doit être incluse dans tout système de stockage prenant en charge les charges de travail de la base de données.

Panne de disque : RAID, RAID DP et RAID-TEC

Si un bloc de données sur un disque est détecté comme étant corrompu, ou si l'ensemble du disque tombe en panne et est totalement indisponible, les données doivent être reconstituées. Cette opération est réalisée dans ONTAP à l'aide de disques de parité. Les données sont réparties sur plusieurs disques, puis des données de parité sont générées. Ces données sont stockées séparément des données d'origine.

ONTAP utilisait à l'origine RAID 4, qui utilise un seul lecteur de parité pour chaque groupe de lecteurs de données. Le résultat a été qu'un disque du groupe pouvait tomber en panne sans entraîner de perte de données. En cas de panne du disque de parité, aucune donnée n'a été endommagée et un nouveau disque de parité a pu être construit. En cas de panne d'un seul lecteur de données, les lecteurs restants peuvent être utilisés avec le lecteur de parité pour régénérer les données manquantes.

Lorsque les disques étaient petits, le risque statistique de défaillance simultanée de deux disques était négligeable. Avec l'augmentation des capacités des disques, la reconstruction des données suite à une panne disque s'est également accompagnée d'un temps considérable. Cela a augmenté la fenêtre au cours de laquelle une panne de second disque entraînerait la perte de données. De plus, le processus de reconstruction crée une grande quantité d'E/S supplémentaires sur les disques survivants. Au fur et à mesure du vieillissement des disques, le risque d'une charge supplémentaire entraînant une panne de second disque augmente également. Enfin, même si le risque de perte de données n'augmente pas avec l'utilisation continue de RAID 4, les conséquences de la perte de données deviendront plus graves. Plus la perte de données en cas de panne d'un groupe RAID est importante, plus la restauration des données est longue, ce qui entraîne une interruption de l'activité prolongée.

Ces problèmes ont conduit NetApp à développer la technologie NetApp RAID DP, une variante de RAID 6. Cette solution comprend deux disques de parité, ce qui signifie que deux disques d'un groupe RAID peuvent tomber en panne sans générer de perte de données. Les disques ont continué de croître en taille, ce qui a conduit NetApp à développer la technologie NetApp RAID-TEC, qui introduit un troisième disque de parité.

Certaines meilleures pratiques en matière de bases de données historiques recommandent l'utilisation de RAID-10, également appelée mise en miroir par bandes. Cela offre une protection des données inférieure à celle de RAID DP, car il existe plusieurs scénarios de défaillance de deux disques, alors que dans RAID DP, il n'en existe aucune.

Par ailleurs, certaines bonnes pratiques en matière d'historique de bases de données indiquent que RAID-10 est préféré aux options RAID-4/5/6 en raison de problèmes de performances. Ces recommandations font parfois référence à une pénalité RAID. Bien que ces recommandations soient généralement correctes, elles ne s'appliquent pas aux implémentations de RAID dans ONTAP. Le problème de performances est lié à la régénération de parité. Dans les implémentations RAID traditionnelles, le traitement des écritures aléatoires de routine effectuées par une base de données nécessite plusieurs lectures de disque pour régénérer les données de parité et terminer l'écriture. La pénalité est définie comme les IOPS de lecture supplémentaires requises pour exécuter les opérations d'écriture.

ONTAP n'engendre pas de pénalité RAID, car les écritures sont placées dans la mémoire où la parité est générée, puis écrites sur le disque sous la forme d'une seule bande RAID. Aucune lecture n'est requise pour terminer l'opération d'écriture.

En résumé, par rapport à RAID 10, les systèmes RAID DP et RAID-TEC fournissent une capacité utilisable nettement plus importante, une meilleure protection contre les pannes disque et sans sacrifier les performances.

Protection contre les pannes matérielles : NVRAM

Toute baie de stockage servant de charge de travail de base de données doit traiter les opérations d'écriture le plus rapidement possible. En outre, une opération d'écriture doit être protégée contre la perte d'un événement inattendu tel qu'une coupure de courant. Cela signifie que toute opération d'écriture doit être stockée en toute sécurité dans au moins deux emplacements.

Les systèmes AFF et FAS utilisent la mémoire NVRAM pour répondre à ces exigences. Le processus d'écriture fonctionne comme suit :

1. Les données d'écriture entrantes sont stockées dans la mémoire RAM.
2. Les modifications à apporter aux données du disque sont journalisées dans la mémoire NVRAM sur le nœud local et le nœud partenaire. La mémoire NVRAM n'est pas un cache d'écriture. Il s'agit plutôt d'un journal similaire à un redo log de base de données. Dans des conditions normales, il n'est pas lu. Il est utilisé uniquement pour la restauration, par exemple après une coupure de courant pendant le traitement des E/S.
3. L'écriture est alors validée par l'hôte.

À ce stade, le processus d'écriture est complet du point de vue de l'application. Les données sont protégées contre les pertes, car elles sont stockées dans deux emplacements différents. Finalement, les modifications sont écrites sur le disque, mais ce processus est hors bande du point de vue de l'application, car il se produit après l'acquittement de l'écriture et n'affecte donc pas la latence. Ce processus est une fois de plus similaire à la journalisation de la base de données. Une modification de la base de données est enregistrée dans les journaux de reprise aussi rapidement que possible, et la modification est alors reconnue comme validée. Les mises à jour des fichiers de données sont effectuées beaucoup plus tard et n'affectent pas directement la vitesse de traitement.

En cas de panne de contrôleur, le contrôleur partenaire prend possession des disques requis et lit à nouveau les données consignées dans la mémoire NVRAM pour récupérer toutes les opérations d'E/S en cours de fonctionnement au moment de la défaillance.

Protection contre les défaillances matérielles : NVFAIL

Comme nous l'avons vu précédemment, une écriture n'est pas validée tant qu'elle n'a pas été connectée à la NVRAM et à la NVRAM locales sur au moins un autre contrôleur. Cette approche évite toute panne matérielle ou de courant qui entraîne une perte des E/S à la volée En cas de panne de la mémoire NVRAM locale ou de la connectivité au partenaire de haute disponibilité, ces données à la volée ne seront plus mises en miroir.

Si la mémoire NVRAM locale signale une erreur, le nœud s'arrête. Cet arrêt entraîne le basculement vers un contrôleur partenaire de haute disponibilité. Aucune donnée n'est perdue parce que le contrôleur qui connaît la défaillance n'a pas acquitté l'opération d'écriture.

ONTAP n'autorise pas le basculement lorsque les données sont désynchronisées, sauf si le basculement est forcé. Le fait de forcer une modification des conditions de cette manière reconnaît que les données peuvent être laissées pour compte dans le contrôleur d'origine et que la perte de données est acceptable.

Les bases de données sont particulièrement vulnérables à la corruption en cas de basculement forcé, car elles conservent de grands caches internes de données sur disque. En cas de basculement forcé, les modifications précédemment reconnues sont effectivement supprimées. Le contenu de la baie de stockage recule dans le temps et l'état du cache de la base de données ne reflète plus l'état des données sur le disque.

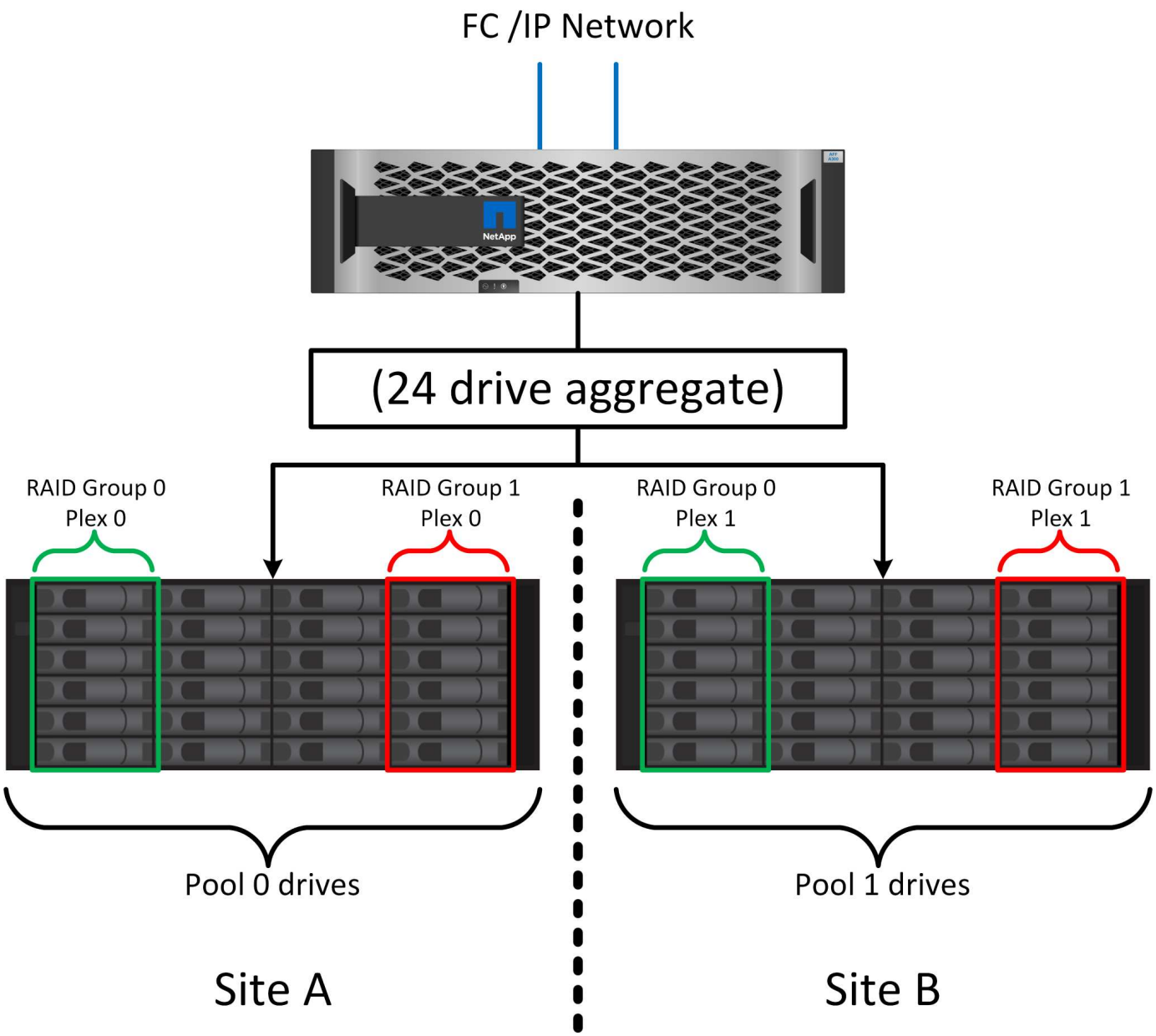
Afin de protéger les données de cette situation, ONTAP permet de configurer les volumes pour une protection spéciale contre les défaillances de mémoire NVRAM. Lorsqu'il est déclenché, ce mécanisme de protection entraîne l'entrée d'un volume dans un état appelé NVFAIL. Cet état entraîne des erreurs d'E/S qui entraînent l'arrêt d'une application et n'utilisent donc pas de données obsolètes. Les données ne doivent pas être perdues car une écriture reconnue doit être présente sur la matrice de stockage.

Les étapes suivantes habituelles sont qu'un administrateur arrête complètement les hôtes avant de remettre manuellement en ligne les LUN et les volumes. Bien que ces étapes puissent impliquer un certain travail, cette approche est le moyen le plus sûr d'assurer l'intégrité des données. Toutes les données n'ont pas besoin de cette protection. C'est pourquoi NVFAIL peut être configuré volume par volume.

Protection contre les pannes de site et de tiroir : SyncMirror et plexes

SyncMirror est une technologie de mise en miroir qui améliore, mais ne remplace pas, RAID DP ou RAID-TEC. Il met en miroir le contenu de deux groupes RAID indépendants. La configuration logique est la suivante :

- Les disques sont configurés en deux pools en fonction de leur emplacement. Un pool est composé de tous les disques du site A et le second est composé de tous les disques du site B.
- Un pool de stockage commun, appelé agrégat, est ensuite créé à partir de jeux en miroir de groupes RAID. Un nombre égal de lecteurs est tiré de chaque site. Par exemple, un agrégat SyncMirror de 20 disques se compose de 10 disques du site A et de 10 disques du site B.
- Chaque jeu de disques d'un site donné est automatiquement configuré comme un ou plusieurs groupes RAID-DP ou RAID-TEC entièrement redondants, indépendamment de l'utilisation de la mise en miroir. Les données sont ainsi protégées en permanence, même après la perte d'un site.



La figure ci-dessus illustre un exemple de configuration SyncMirror. Un agrégat de 24 disques a été créé sur le contrôleur avec 12 disques à partir d'un tiroir alloué sur le site A et 12 disques à partir d'un tiroir alloué sur le site B. Les disques ont été regroupés en deux groupes RAID en miroir. Le groupe RAID 0 comprend un plex de 6 disques sur le site A mis en miroir sur un plex de 6 disques sur le site B. De même, RAID Group 1 inclut un plex de 6 disques sur le site A mis en miroir sur un plex de 6 disques sur le site B.

SyncMirror est généralement utilisé pour assurer la mise en miroir à distance avec les systèmes MetroCluster, avec une copie des données sur chaque site. Il a parfois été utilisé pour fournir un niveau supplémentaire de redondance dans un seul système. Il assure en particulier la redondance au niveau du tiroir. Un tiroir disque contient déjà deux blocs d'alimentation et contrôleurs. Dans l'ensemble, il ne s'agit pas d'une simple tôlerie, mais dans certains cas, une protection supplémentaire peut être garantie. Par exemple, un client NetApp a déployé SyncMirror sur une plateforme mobile d'analytique en temps réel utilisée lors des tests automobiles. Le système a été séparé en deux racks physiques alimentés par des alimentations indépendantes provenant de systèmes UPS indépendants.

Checksums

Le thème des checksums est particulièrement intéressant pour les administrateurs de bases de données habitués à l'utilisation de sauvegardes en continu Oracle RMAN qui migrent vers des sauvegardes basées sur des snapshots. RMAN permet notamment de procéder à des contrôles d'intégrité lors des opérations de sauvegarde. Bien que cette fonctionnalité présente un certain intérêt, son principal avantage est une base de données qui n'est pas utilisée sur une baie de stockage moderne. Lorsque des disques physiques sont utilisés pour une base de données Oracle, il est presque certain que la corruption finit par se produire lorsque les disques vieillissent, un problème qui est résolu par les checksums basés sur les baies dans les baies de stockage réelles.

Avec une baie de stockage réelle, l'intégrité des données est protégée par des checksums à plusieurs niveaux. Si les données sont corrompues dans un réseau IP, la couche TCP (transmission Control Protocol) rejette les données de paquets et demande la retransmission. Le protocole FC inclut des checksums, tout comme les données SCSI encapsulées. Une fois sur la matrice, ONTAP dispose d'une protection RAID et checksum. Une corruption peut se produire, mais, comme dans la plupart des baies d'entreprise, elle est détectée et corrigée. En général, un disque entier tombe en panne, ce qui invite à une reconstruction RAID et l'intégrité de la base de données n'est pas affectée. Il est toujours possible que des octets individuels sur un disque soient endommagés par le rayonnement cosmique ou par des cellules flash défectueuses. Si cela se produit, la vérification de parité échoue, le disque est mis hors service et la reconstruction RAID démarre. Là encore, l'intégrité des données n'est pas affectée. La dernière ligne de défense est l'utilisation de checksums. Si, par exemple, une erreur de micrologiciel catastrophique sur un disque a corrompu des données d'une manière qui n'a pas été détectée par un contrôle de parité RAID, le checksum ne correspond pas et ONTAP empêche le transfert d'un bloc corrompu avant que la base de données Oracle puisse le recevoir.

L'architecture des fichiers de données et des redo log Oracle est également conçue pour offrir le plus haut niveau possible d'intégrité des données, même dans des circonstances extrêmes. Au niveau le plus élémentaire, les blocs Oracle incluent un checksum et des contrôles logiques de base avec presque toutes les E/S. Si Oracle ne s'est pas écrasé ou n'a pas mis un tablespace hors ligne, les données sont intactes. Le degré de vérification de l'intégrité des données est réglable et Oracle peut également être configuré pour confirmer les écritures. Par conséquent, la quasi-totalité des scénarios de panne et de panne peuvent être restaurés, et dans le cas extrêmement rare d'une situation irrécupérable, la corruption est rapidement détectée.

La plupart des clients NetApp qui utilisent des bases de données Oracle cessent d'utiliser RMAN et d'autres produits de sauvegarde après la migration vers des sauvegardes snapshot. Il existe encore des options permettant d'utiliser RMAN pour effectuer une restauration au niveau des blocs avec SnapCenter. Toutefois, au quotidien, RMAN, NetBackup et d'autres produits ne sont utilisés qu'occasionnellement pour créer des copies d'archivage mensuelles ou trimestrielles.

Certains clients choisissent d'exécuter `dbv` périodiquement pour effectuer des contrôles d'intégrité sur leurs bases de données existantes. NetApp déconseille cette pratique, car elle entraîne une charge d'E/S inutile. Comme indiqué ci-dessus, si la base de données ne rencontrait pas de problèmes auparavant, le risque de `dbv` La détection d'un problème est proche de zéro et cet utilitaire entraîne une charge d'E/S séquentielles très élevée sur le réseau et le système de stockage. À moins qu'il n'y ait de raison de croire qu'il existe une corruption, comme l'exposition à un bogue connu d'Oracle, il n'y a aucune raison de s'exécuter `dbv`.

Notions de base sur la sauvegarde et la restauration

Sauvegardes basées sur des snapshots

La technologie Snapshot de NetApp constitue le socle de la protection des données des bases de données Oracle sur ONTAP.

Les valeurs clés sont les suivantes :

- **Simplicité.** Un instantané est une copie en lecture seule du contenu d'un conteneur de données à un moment donné.
- **Efficacité.** les instantanés ne nécessitent pas d'espace au moment de la création. L'espace n'est consommé que lorsque des données sont modifiées.
- **Gérabilité.** Une stratégie de sauvegarde basée sur les snapshots est facile à configurer et à gérer car les snapshots font partie intégrante du système d'exploitation du stockage. Si le système de stockage est sous tension, il est prêt à créer des sauvegardes.
- **Évolutivité.** vous pouvez conserver jusqu'à 1024 sauvegardes d'un seul conteneur de fichiers et de LUN. Dans le cas de jeux de données complexes, plusieurs conteneurs de données peuvent être protégés par un ensemble unique et cohérent de snapshots.
- Les performances ne sont pas affectées, qu'un volume contienne ou non 1024 snapshots.

Bien que de nombreux fournisseurs de stockage proposent la technologie Snapshot, la technologie Snapshot de ONTAP est unique et offre des avantages significatifs pour les environnements applicatifs et de bases de données d'entreprise :

- Les copies Snapshot font partie de la WAFL (Write-Anywhere File Layout) sous-jacente. Il ne s'agit pas d'une technologie complémentaire ou externe. La gestion est donc simplifiée, car le système de stockage est le système de sauvegarde.
- Les copies Snapshot n'affectent pas les performances, sauf dans certains cas en périphérie, par exemple lorsque le volume de données est stocké dans des snapshots que le système de stockage sous-jacent se remplit.
- Le terme « groupe de cohérence » fait souvent référence à un regroupement d'objets de stockage gérés comme un ensemble cohérent de données. La copie Snapshot d'un volume ONTAP donné constitue une sauvegarde de groupe de cohérence.

Les copies Snapshot ONTAP ont également une meilleure évolutivité que la technologie concurrente. Les clients peuvent stocker 5, 50 ou 500 copies Snapshot sans affecter les performances. Le nombre maximal de snapshots actuellement autorisés dans un volume est de 1024. Si une conservation supplémentaire des snapshots est nécessaire, il existe des options pour les transmettre en cascade à des volumes supplémentaires.

Par conséquent, la protection d'un dataset hébergé sur ONTAP est simple et hautement évolutive. Les sauvegardes ne nécessitent pas de déplacement de données. Par conséquent, une stratégie de sauvegarde peut être adaptée aux besoins de l'entreprise plutôt qu'aux limites des taux de transfert réseau, du grand nombre de lecteurs de bande ou des zones de transfert de disque.

Un snapshot est-il une sauvegarde ?

La question couramment posée sur l'utilisation des snapshots en tant que stratégie de protection des données est le fait que les données « réelles » et les données de snapshot se trouvent sur les mêmes disques. La perte de ces disques entraînerait la perte des données primaires et de la sauvegarde.

Ce problème est valide. Les snapshots locaux sont utilisés pour les besoins quotidiens de sauvegarde et de restauration, et dans ce sens, le snapshot est une sauvegarde. Dans les environnements NetApp, près de 99 % des scénarios de restauration s'appuient sur des copies Snapshot pour répondre aux exigences de RTO les plus strictes.

Toutefois, les snapshots locaux ne doivent jamais être la seule stratégie de sauvegarde. C'est pourquoi NetApp propose des technologies telles que la réplication SnapMirror et SnapVault pour répliquer rapidement

et efficacement des copies Snapshot sur un ensemble indépendant de disques. Dans une solution bien conçue avec des snapshots et une réplication Snapshot, l'utilisation des bandes peut être réduite au minimum, voire même à une archive trimestrielle, ou totalement éliminée.

Sauvegardes basées sur des snapshots

Vous pouvez utiliser les copies Snapshot ONTAP pour protéger vos données, et les copies Snapshot sont la base de nombreuses autres fonctionnalités ONTAP, notamment la réplication, la reprise d'activité et le clonage. Une description complète de la technologie Snapshot ne fait pas partie du présent document, mais les sections suivantes offrent un aperçu général.

Il existe deux approches principales pour créer un snapshot d'un dataset :

- Sauvegardes cohérentes après panne
- Sauvegardes cohérentes au niveau des applications

Une sauvegarde cohérente après panne d'un dataset fait référence à la capture de l'ensemble de la structure du dataset à un point dans le temps. Si le dataset est stocké dans un seul volume, le processus est simple ; il est possible de créer une copie Snapshot à tout moment. Si un dataset s'étend sur plusieurs volumes, un snapshot de groupe de cohérence doit être créé. Plusieurs options sont disponibles pour la création des snapshots de groupe de cohérence, notamment le logiciel NetApp SnapCenter, les fonctionnalités natives de groupe de cohérence ONTAP et les scripts gérés par l'utilisateur.

Les sauvegardes cohérentes après panne sont principalement utilisées lorsque la restauration au point de sauvegarde est suffisante. Lorsqu'une restauration plus granulaire est nécessaire, des sauvegardes cohérentes au niveau des applications sont généralement nécessaires.

Le mot "cohérent" dans "application-cohérente" est souvent un mal nommer. Par exemple, le placement d'une base de données Oracle en mode de sauvegarde est appelé sauvegarde cohérente au niveau des applications, mais les données ne sont en aucun cas rendues cohérentes ou suspendues. Les données continuent de changer tout au long de la sauvegarde. En revanche, la plupart des sauvegardes MySQL et Microsoft SQL Server ont effectivement mis les données au repos avant d'exécuter la sauvegarde. VMware peut rendre certains fichiers cohérents ou non.

Groupes de cohérence

Le terme « groupe de cohérence » fait référence à la capacité d'une baie de stockage à gérer plusieurs ressources de stockage comme une seule image. Par exemple, une base de données peut comprendre 10 LUN. La baie doit pouvoir sauvegarder, restaurer et répliquer ces 10 LUN de manière cohérente. La restauration n'est pas possible si les images des LUN n'étaient pas cohérentes au point de sauvegarde. La réplication de ces 10 LUN nécessite que tous les réplicas soient parfaitement synchronisés.

Le terme « groupe de cohérence » n'est pas souvent utilisé lors des discussions sur ONTAP, car la cohérence a toujours été une fonction de base de l'architecture de volumes et d'agrégats au sein de ONTAP. De nombreuses autres baies de stockage gèrent des LUN ou des systèmes de fichiers en tant qu'unités individuelles. Ils peuvent ensuite être configurés en tant que « groupe de cohérence » pour la protection des données, mais cette étape supplémentaire est nécessaire dans la configuration.

ONTAP a toujours pu capturer des images locales et répliquées cohérentes de données. Bien que les différents volumes d'un système ONTAP ne soient généralement pas officiellement décrits comme des groupes de cohérence, c'est ce qu'ils sont. Une copie Snapshot de ce volume est une image de groupe de cohérence. La restauration de ce Snapshot correspond à une restauration de groupe de cohérence. SnapMirror et SnapVault proposent tous deux une réplication de groupe de cohérence.

Snapshots de groupes de cohérence

Les copies Snapshot de groupe de cohérence (cg-snapshots) sont une extension de la technologie Snapshot ONTAP de base. Une opération de snapshot standard crée une image cohérente de toutes les données d'un même volume, mais il est parfois nécessaire de créer un ensemble cohérent de snapshots sur plusieurs volumes et même sur plusieurs systèmes de stockage. Il en résulte un ensemble de snapshots qui peuvent être utilisés de la même manière qu'un snapshot d'un seul volume individuel. Elles peuvent être utilisées pour la restauration des données locales, répliquées à des fins de reprise après incident ou clonées sous la forme d'une unité cohérente unique.

L'utilisation la plus connue des cg-snapshots concerne un environnement de base de données d'environ 1 po de capacité couvrant 12 contrôleurs. Les snapshots de groupe de cohérence créés sur ce système ont été utilisés pour la sauvegarde, la restauration et le clonage.

La plupart du temps, lorsqu'un dataset s'étend sur des volumes et que l'ordre d'écriture doit être préservé, le logiciel de gestion choisi utilise automatiquement un snapshot de groupe de cohérence. Dans ce cas, il n'est pas nécessaire de comprendre les détails techniques des cg-snapshots. Toutefois, les exigences complexes en matière de protection des données nécessitent un contrôle détaillé du processus de protection et de réplication des données. Certains workflows d'automatisation ou scripts personnalisés permettent d'appeler les API cg-Snapshot. Pour comprendre la meilleure option et le rôle de cg-snapshot, vous devez fournir une explication plus détaillée de la technologie.

La création d'un ensemble de snapshots des groupes de cohérence s'effectue en deux étapes :

1. Établir une clôture d'écriture sur tous les volumes cibles.
2. Créez des instantanés de ces volumes à l'état clôturé.

L'escrime d'écriture est établi en série. Cela signifie que lorsque le processus de recel est configuré sur plusieurs volumes, les E/S d'écriture sont bloquées sur le premier volume de la séquence au fur et à mesure qu'elles continuent d'être validées sur les volumes qui apparaissent plus tard. Cela peut sembler initialement contraire à l'exigence de préservation de l'ordre d'écriture, mais cela s'applique uniquement aux E/S émises de manière asynchrone sur l'hôte et ne dépend pas d'autres écritures.

Par exemple, une base de données peut émettre de nombreuses mises à jour asynchrones des fichiers de données et permettre au système d'exploitation de réorganiser les E/S et de les compléter selon sa propre configuration de planificateur. L'ordre de ce type d'E/S ne peut pas être garanti car l'application et le système d'exploitation ont déjà libéré l'obligation de conserver l'ordre d'écriture.

Par exemple, la plupart des activités de journalisation de la base de données sont synchrones. La base de données ne procède pas à d'autres écritures de journal tant que les E/S n'ont pas été acquittées et que l'ordre de ces écritures doit être conservé. Si une E/S de journal arrive sur un volume clôturé, elle n'est pas validée et l'application se bloque lors d'écritures ultérieures. De même, les E/S des métadonnées du système de fichiers sont généralement synchrones. Par exemple, une opération de suppression de fichier ne doit pas être perdue. Si un système d'exploitation doté d'un système de fichiers xfs supprime un fichier et que les E/S qui ont mis à jour les métadonnées du système de fichiers xfs pour supprimer la référence à ce fichier ont été reçues sur un volume isolé, l'activité du système de fichiers est alors interrompue. Cela garantit l'intégrité du système de fichiers pendant les opérations cg-Snapshot.

Une fois l'isolation d'écriture configurée sur les volumes cibles, ils sont prêts pour la création d'instantanés. Les snapshots n'ont pas besoin d'être créés précisément en même temps, car l'état des volumes est figé du point de vue de l'écriture dépendant. Pour éviter toute faille dans l'application qui crée les instantanés cg, l'escrime d'écriture initiale inclut un délai configurable dans lequel ONTAP libère automatiquement l'escrime et reprend le traitement d'écriture après un nombre défini de secondes. Si tous les snapshots sont créés avant l'expiration du délai, le jeu de snapshots résultant est un groupe de cohérence valide.

Ordre d'écriture dépendant

Du point de vue technique, la préservation de l'ordre d'écriture et, plus particulièrement, de l'ordre d'écriture dépendant constitue la clé d'un groupe de cohérence. Par exemple, une base de données qui écrit 10 LUN écrit simultanément sur toutes ces LUN. De nombreuses écritures sont émises de manière asynchrone, ce qui signifie que l'ordre dans lequel elles sont effectuées n'est pas important et que l'ordre dans lequel elles sont effectuées varie en fonction du système d'exploitation et du comportement du réseau.

Certaines opérations d'écriture doivent être présentes sur le disque avant que la base de données puisse procéder à des écritures supplémentaires. Ces opérations d'écriture critiques sont appelées écritures dépendantes. Les E/S d'écriture suivantes dépendent de la présence de ces écritures sur le disque. Tout snapshot, restauration ou réplication de ces 10 LUN doit garantir l'ordre d'écriture dépendant. Les mises à jour du système de fichiers sont un autre exemple d'écritures dépendantes de l'ordre d'écriture. L'ordre dans lequel les modifications du système de fichiers sont effectuées doit être conservé, sinon l'ensemble du système de fichiers pourrait être corrompu.

Stratégies

Il existe deux approches principales des sauvegardes basées sur des snapshots :

- Sauvegardes cohérentes après panne
- Sauvegardes à chaud protégées pour les snapshots

Une sauvegarde cohérente après panne d'une base de données fait référence à la capture à un moment précis de l'ensemble de la structure de la base de données, y compris les fichiers de données, les journaux de reprise et les fichiers de contrôle. Si la base de données est stockée sur un seul volume, le processus est simple ; il est possible de créer un Snapshot à tout moment. Si la base de données s'étend sur plusieurs volumes, un snapshot de groupe de cohérence doit être créé. Plusieurs options sont disponibles pour la création des snapshots de groupe de cohérence, notamment le logiciel NetApp SnapCenter, les fonctionnalités natives de groupe de cohérence ONTAP et les scripts gérés par l'utilisateur.

Les sauvegardes Snapshot cohérentes après panne sont principalement utilisées lorsque la restauration au point de sauvegarde est suffisante. Les journaux d'archivage peuvent être appliqués dans certains cas, mais lorsqu'une restauration granulaire à un point dans le temps est nécessaire, il est préférable d'effectuer une sauvegarde en ligne.

La procédure de base pour une sauvegarde en ligne basée sur un snapshot est la suivante :

1. Placez la base de données dans `backup` mode.
2. Créez un Snapshot de tous les volumes qui hébergent les fichiers de données.
3. Quitter `backup` mode.
4. Lancer la commande `alter system archive log current` pour forcer l'archivage des journaux.
5. Créer des instantanés de tous les volumes hébergeant les journaux d'archivage.

Cette procédure permet d'obtenir un ensemble de snapshots contenant les fichiers de données en mode de sauvegarde et les journaux d'archivage critiques générés en mode de sauvegarde. Il s'agit des deux conditions requises pour restaurer une base de données. Il est également conseillé de protéger les fichiers tels que les fichiers de contrôle, mais la seule condition absolue est la protection des fichiers de données et des journaux d'archivage.

Même si différents clients peuvent avoir des stratégies très différentes, la quasi-totalité de ces stratégies s'appuient sur les mêmes principes que ceux décrits ci-dessous.

Restauration basée sur des snapshots

Lors de la conception d'infrastructures de volumes pour les bases de données Oracle, la première décision est d'utiliser ou non la technologie VBSR (Volume-Based NetApp SnapRestore).

La fonction SnapRestore basée sur les volumes permet de rétablir quasi instantanément un volume à un point antérieur. Toutes les données du volume étant rétablies, VBSR peut ne pas convenir à toutes les utilisations. Par exemple, si l'intégralité d'une base de données, y compris les fichiers de données, les journaux de reprise et les journaux d'archivage, est stockée sur un seul volume restauré avec VBSR, les données sont perdues, car les nouveaux journaux d'archivage et les données de reprise sont supprimés.

La technologie VBSR n'est pas requise pour la restauration. De nombreuses bases de données peuvent être restaurées avec SFSR (Single File SnapRestore) ou en copiant simplement les fichiers du snapshot vers le système de fichiers actif.

La technologie VBSR est recommandée pour les bases de données très volumineuses ou si une restauration doit être effectuée le plus rapidement possible et que l'utilisation de VBSR nécessite l'isolement des fichiers de données. Dans un environnement NFS, les fichiers de données d'une base de données doivent être stockés sur des volumes dédiés non endommagés par d'autres types de fichiers. Dans un environnement SAN, les fichiers de données doivent être stockés sur des LUN dédiés sur des volumes dédiés. Si un gestionnaire de volumes est utilisé (y compris Oracle Automatic Storage Management (ASM)), le groupe de disques doit également être dédié aux fichiers de données.

Cette méthode d'isolement des fichiers de données permet de rétablir leur état antérieur sans endommager d'autres systèmes de fichiers.

Réserve Snapshot

Pour chaque volume contenant des données Oracle dans un environnement SAN, le `percent-snapshot-space` Doit être défini sur zéro car il n'est pas utile de réserver de l'espace pour un snapshot dans un environnement LUN. Si la réserve fractionnaire est définie sur 100, un snapshot d'un volume avec des LUN nécessite suffisamment d'espace libre dans le volume, à l'exception de la réserve Snapshot, pour absorber 100 % de CA de toutes les données. Si la réserve fractionnaire est définie sur une valeur inférieure, une quantité d'espace libre correspondante est nécessaire, mais elle exclut toujours la réserve snapshot. Cela signifie que l'espace de réserve du snapshot dans un environnement de LUN est gaspillé.

Dans un environnement NFS, deux options sont possibles :

- Réglez le `percent-snapshot-space` basé sur la consommation d'espace prévue du snapshot.
- Réglez le `percent-snapshot-space` pour zéro et gérer collectivement l'espace utilisé actif et snapshot.

Avec la première option, `percent-snapshot-space` est défini sur une valeur différente de zéro, généralement autour de 20 %. Cet espace est alors masqué par l'utilisateur. Toutefois, cette valeur ne crée pas de limite d'utilisation. Si une base de données avec une réservation de 20 % connaît un chiffre d'affaires de 30 %, l'espace snapshot peut dépasser les limites de la réserve de 20 % et occuper un espace non réservé.

Le principal avantage de la définition d'une réserve sur une valeur telle que 20 % est de vérifier qu'un peu d'espace est toujours disponible pour les snapshots. Par exemple, un volume de 1 To avec une réserve de 20 % permettrait uniquement à un administrateur de base de données (DBA) de stocker 800 Go de données. Cette configuration garantit au moins 200 Go d'espace pour la consommation de snapshots.

Quand `percent-snapshot-space` est défini sur zéro, tout l'espace du volume est disponible pour l'utilisateur final, ce qui offre une meilleure visibilité. L'administrateur de base de données doit comprendre que,

s'il constate qu'un volume de 1 To exploite les snapshots, cet espace de 1 To est partagé entre les données actives et le renouvellement du Snapshot.

Il n'existe pas de préférence claire entre l'option 1 et l'option 2 parmi les utilisateurs finaux.

ONTAP et snapshots tiers

Oracle Doc ID 604683.1 décrit les conditions requises pour la prise en charge des snapshots tiers et les nombreuses options disponibles pour les opérations de sauvegarde et de restauration.

Les fournisseurs tiers doivent garantir la conformité de leurs snapshots à plusieurs exigences :

- Les snapshots doivent intégrer les opérations de restauration et de reprise recommandées par Oracle.
- Les snapshots doivent être cohérents après panne de la base de données au point du Snapshot.
- L'ordre d'écriture est conservé pour chaque fichier d'un snapshot.

Les produits de gestion Oracle de ONTAP et NetApp sont conformes à ces exigences.

SnapRestore

La technologie NetApp SnapRestore assure la restauration rapide des données dans ONTAP à partir d'une copie Snapshot.

Lorsqu'un dataset stratégique n'est pas disponible, les opérations stratégiques de l'entreprise ne sont pas disponibles. Les bandes peuvent se rompre, et même les restaurations à partir de sauvegardes sur disque peuvent être lentes à transférer sur le réseau. SnapRestore évite ces problèmes en offrant une restauration quasi instantanée des datasets. Même les bases de données de plusieurs pétaoctets peuvent être entièrement restaurées en quelques minutes à peine.

Il existe deux types d'SnapRestore : basés sur les fichiers/LUN et sur les volumes.

- Il est possible de restaurer des fichiers individuels ou des LUN en quelques secondes, qu'il s'agisse d'un LUN de 2 To ou d'un fichier de 4 Ko.
- Le conteneur de fichiers ou de LUN peut être restauré en quelques secondes, qu'il s'agisse de 10 Go ou 100 To de données.

Un « conteneur de fichiers ou de LUN » fait généralement référence à un volume FlexVol. Par exemple, vous pouvez avoir 10 LUN qui composent un groupe de disques LVM dans un seul volume, ou un volume peut stocker les home directories NFS de 1000 utilisateurs. Au lieu d'exécuter une opération de restauration pour chaque fichier ou LUN individuel, vous pouvez restaurer le volume entier en une seule opération. Ce processus fonctionne également avec des conteneurs scale-out qui incluent plusieurs volumes, tels qu'un FlexGroup ou un groupe de cohérence ONTAP.

La rapidité et l'efficacité de SnapRestore sont dues à la nature d'une copie Snapshot, qui offre essentiellement une vue en lecture seule parallèle du contenu d'un volume à un moment donné. Les blocs actifs sont les blocs réels qui peuvent être modifiés, tandis que le snapshot offre une vue en lecture seule de l'état des blocs qui constituent les fichiers et les LUN au moment de la création du snapshot.

ONTAP permet uniquement un accès en lecture seule aux données instantanées, mais les données peuvent être réactivées avec SnapRestore. L'instantané est réactivé en tant que vue en lecture-écriture des données, renvoyant les données à leur état précédent. SnapRestore peut fonctionner au niveau du volume ou du fichier. La technologie est essentiellement la même avec quelques différences mineures de comportement.

SnapRestore du volume

La fonction SnapRestore basée sur les volumes renvoie la totalité du volume de données à un état antérieur. Cette opération ne nécessite pas de déplacement de données. Le processus de restauration est donc pratiquement instantané, bien que le traitement des opérations via l'API ou l'interface de ligne de commande puisse prendre quelques secondes. La restauration de 1 Go de données n'est pas plus compliquée et chronophage que la restauration de 1 po de données. Cette fonctionnalité est la principale raison pour laquelle de nombreux clients grands comptes migrent vers des systèmes de stockage ONTAP. Il assure un RTO se mesure en quelques secondes, même pour les datasets les plus volumineux.

L'un des inconvénients des SnapRestore sur volume est le fait que les modifications au sein d'un volume sont cumulées dans le temps. Par conséquent, chaque snapshot et les données de fichier actives dépendent des modifications apportées jusqu'à ce point. Le rétablissement d'un volume à un état antérieur implique la suppression de toutes les modifications ultérieures apportées aux données. Ce qui est moins évident, cependant, c'est qu'il s'agit d'instantanés créés par la suite. Ce n'est pas toujours souhaitable.

Par exemple, un SLA de conservation des données peut spécifier 30 jours de sauvegardes nocturnes. La restauration d'un dataset sur un snapshot créé il y a cinq jours avec SnapRestore du volume abandonnerait tous les snapshots créés les cinq jours précédents, en violation du SLA.

Un certain nombre d'options sont disponibles pour résoudre cette limitation :

1. Les données peuvent être copiées à partir d'un instantané précédent, au lieu d'effectuer une SnapRestore du volume entier. Cette méthode fonctionne mieux avec les jeux de données plus petits.
2. Un snapshot peut être cloné plutôt que restauré. La limitation à cette approche est que le snapshot source dépend du clone. Par conséquent, elle ne peut pas être supprimée si le clone n'est pas également supprimé ou s'il est divisé en volume indépendant.
3. Utilisation d'un SnapRestore basé sur des fichiers.

Fichier SnapRestore

SnapRestore basé sur les fichiers est un processus de restauration plus granulaire basé sur des snapshots. Au lieu de rétablir l'état d'un volume entier, l'état d'un fichier ou d'une LUN individuel est rétabli. Il n'est pas nécessaire de supprimer des snapshots et cette opération ne crée aucune dépendance vis-à-vis d'un instantané précédent. Le fichier ou la LUN est immédiatement disponible dans le volume actif.

Aucun déplacement des données n'est nécessaire lors de la restauration d'un fichier ou d'une LUN par SnapRestore. Cependant, des mises à jour internes des métadonnées sont nécessaires pour refléter le fait que les blocs sous-jacents d'un fichier ou d'une LUN existent désormais à la fois dans un snapshot et dans le volume actif. Les performances ne doivent pas être affectées, mais ce processus bloque la création de snapshots jusqu'à ce qu'elle soit terminée. Le taux de traitement est d'environ 5 Gbit/s (18 To/heure) en fonction de la taille totale des fichiers restaurés.

Sauvegardes en ligne

Deux datasets sont nécessaires pour protéger et restaurer une base de données Oracle en mode de sauvegarde. Notez qu'il ne s'agit pas de la seule option de sauvegarde Oracle, mais qu'elle est la plus courante.

- Un Snapshot des fichiers de données en mode de sauvegarde
- Les journaux d'archivage créés pendant que les fichiers de données étaient en mode de sauvegarde

Si une récupération complète incluant toutes les transactions validées est requise, un troisième élément est

requis :

- Les journaux de reprise en cours

Il existe plusieurs façons de restaurer une sauvegarde en ligne. De nombreux clients restaurent les snapshots à l'aide de l'interface de ligne de commande ONTAP, puis à l'aide d'Oracle RMAN ou de sqlplus pour terminer la restauration. Cette approche est particulièrement fréquente dans les environnements de production de grande taille. En effet, la probabilité et la fréquence des restaurations de bases de données sont extrêmement faibles et les restaurations sont gérées par un administrateur de bases de données qualifié. Pour une automatisation totale, des solutions telles que NetApp SnapCenter intègrent un plug-in Oracle avec une ligne de commande et des interfaces graphiques.

Certains grands clients ont adopté une approche plus simple en configurant des scripts de base sur les hôtes afin de placer les bases de données en mode de sauvegarde à un moment spécifique en préparation d'un snapshot planifié. Par exemple, planifiez la commande `alter database begin backup` à 23:58, `alter database end backup` à 00:02, puis planifiez les snapshots directement sur le système de stockage à minuit. Résultat : une stratégie de sauvegarde simple et hautement évolutive ne nécessite aucun logiciel ni licence externe.

Disposition des données

La disposition la plus simple consiste à isoler les fichiers de données dans un ou plusieurs volumes dédiés. Ils doivent être non contaminés par tout autre type de fichier. Cela permet de s'assurer que les volumes de fichiers de données peuvent être rapidement restaurés via une opération SnapRestore sans détruire un journal de reprise, un fichier de contrôle ou un journal d'archivage important.

LE SYSTÈME SAN présente des exigences similaires en matière d'isolation des fichiers de données dans des volumes dédiés. Avec un système d'exploitation tel que Microsoft Windows, un seul volume peut contenir plusieurs LUN de fichiers de données, chacune avec un système de fichiers NTFS. Avec d'autres systèmes d'exploitation, il existe généralement un gestionnaire de volumes logiques. Par exemple, avec Oracle ASM, l'option la plus simple consiste à limiter les LUN d'un groupe de disques ASM à un seul volume pouvant être sauvegardé et restauré en tant qu'unité. Si des volumes supplémentaires sont nécessaires pour des raisons de performance ou de gestion de la capacité, la création d'un groupe de disques supplémentaire sur le nouveau volume simplifie la gestion.

Si ces instructions sont respectées, les snapshots peuvent être planifiés directement sur le système de stockage sans avoir à créer de snapshot de groupe de cohérence. En effet, les sauvegardes Oracle ne nécessitent pas la sauvegarde simultanée de fichiers de données. La procédure de sauvegarde en ligne a été conçue pour assurer la mise à jour des fichiers de données, qui seront ensuite transmis progressivement sur bande en quelques heures.

Une complication se produit dans des situations telles que l'utilisation d'un groupe de disques ASM distribué sur des volumes. Dans ce cas, un snapshot de groupe de cohérence doit être réalisé pour s'assurer que les métadonnées ASM sont cohérentes sur tous les volumes constitutifs.

Attention : Vérifiez que l'ASM `spfile` et `passwd` les fichiers ne se trouvent pas dans le groupe de disques hébergeant les fichiers de données. Cela interfère avec la capacité à restaurer de manière sélective les fichiers de données et uniquement les fichiers de données.

Procédure de restauration locale : NFS

Cette procédure peut être conduite manuellement ou via une application telle que SnapCenter. La procédure de base est la suivante :

1. Arrêtez la base de données.

2. Restaurez le ou les volumes de fichiers de données sur l'instantané immédiatement avant le point de restauration souhaité.
3. Réexécutez les journaux d'archivage au point souhaité.
4. Relire les journaux de reprise en cours si vous souhaitez effectuer une restauration complète.

Cette procédure suppose que les journaux d'archive souhaités sont toujours présents dans le système de fichiers actif. Si ce n'est pas le cas, les journaux d'archivage doivent être restaurés ou rman/sqlplus peut être dirigé vers les données du répertoire d'instantanés.

En outre, dans le cas de bases de données plus petites, l'utilisateur peut restaurer les fichiers de données directement à partir du système `.snapshot` répertoire n'ayant pas besoin des outils d'automatisation ou des administrateurs de stockage pour exécuter une `snaprestore` commande.

Procédure de restauration locale—SAN

Cette procédure peut être conduite manuellement ou via une application telle que SnapCenter. La procédure de base est la suivante :

1. Arrêtez la base de données.
2. Arrêter le ou les groupes de disques hébergeant les fichiers de données. La procédure varie en fonction du gestionnaire de volumes logiques choisi. Avec ASM, le processus nécessite de démonter le groupe de disques. Sous Linux, les systèmes de fichiers doivent être démontés et les volumes logiques et les groupes de volumes doivent être désactivés. L'objectif est d'arrêter toutes les mises à jour du groupe de volumes cible à restaurer.
3. Restaurez les groupes de disques de fichiers de données sur l'instantané immédiatement avant le point de restauration souhaité.
4. Réactivez les groupes de disques récemment restaurés.
5. Réexécutez les journaux d'archivage au point souhaité.
6. Relire tous les journaux de reprise si vous souhaitez procéder à une restauration complète.

Cette procédure suppose que les journaux d'archive souhaités sont toujours présents dans le système de fichiers actif. Si ce n'est pas le cas, les journaux d'archivage doivent être restaurés en mettant les LUN du journal d'archivage hors ligne et en effectuant une restauration. Il s'agit également d'un exemple dans lequel il est utile de diviser les journaux d'archivage en volumes dédiés. Si les journaux d'archivage partagent un groupe de volumes avec les journaux de reprise, les journaux de reprise doivent être copiés ailleurs avant la restauration de l'ensemble global des LUN. Cette étape empêche la perte de ces transactions finales enregistrées.

Sauvegardes optimisées pour les snapshots de stockage

La sauvegarde et la restauration basées sur des snapshots sont devenues encore plus simples au moment du lancement d'Oracle 12c. En effet, il n'est pas nécessaire de placer une base de données en mode de sauvegarde à chaud. Il est possible de planifier des sauvegardes Snapshot directement sur un système de stockage et d'effectuer des restaurations complètes ou à un point dans le temps.

Les administrateurs de bases de données maîtrisent mieux la procédure de restauration à partir d'une sauvegarde à chaud, mais il est depuis longtemps possible d'utiliser des snapshots qui n'ont pas été créés pendant que la base de données était en mode de sauvegarde à chaud. Pour assurer la cohérence de la base de données, des étapes manuelles supplémentaires ont été nécessaires avec Oracle 10g et 11g. Avec Oracle

12c, `sqlplus` et `rman` contiennent la logique supplémentaire permettant de relire les journaux d'archivage sur des sauvegardes de fichiers de données qui n'étaient pas en mode de sauvegarde à chaud.

Comme nous l'avons vu précédemment, la restauration d'une sauvegarde à chaud basée sur des snapshots nécessite deux jeux de données :

- Un Snapshot des fichiers de données créés en mode de sauvegarde
- Les journaux d'archivage générés pendant que les fichiers de données étaient en mode de sauvegarde à chaud

Lors de la restauration, la base de données lit les métadonnées à partir des fichiers de données pour sélectionner les journaux d'archivage requis à des fins de restauration.

La restauration optimisée pour les snapshots de stockage nécessite des jeux de données légèrement différents pour obtenir les mêmes résultats :

- Un Snapshot des fichiers de données et une méthode d'identification de l'heure de création du Snapshot
- Archiver les journaux à partir de l'heure du point de contrôle du fichier de données le plus récent jusqu'à l'heure exacte du snapshot

Lors de la restauration, la base de données lit les métadonnées à partir des fichiers de données pour identifier le premier journal d'archivage requis. Il est possible d'effectuer une restauration complète ou instantanée. Lors de l'exécution d'une restauration à un point dans le temps, il est essentiel d'connaître l'heure du Snapshot des fichiers de données. Le point de restauration spécifié doit être après l'heure de création des snapshots. NetApp recommande d'ajouter au moins quelques minutes à l'heure du snapshot pour tenir compte des variations d'horloge.

Pour plus de détails, consultez la documentation d'Oracle sur la rubrique « Restauration à l'aide de l'optimisation des snapshots de stockage » disponible dans les différentes versions de la documentation d'Oracle 12c. Consultez également le document Oracle document ID Doc ID 604683.1 concernant la prise en charge des snapshots tiers par Oracle.

Disposition des données

La disposition la plus simple consiste à isoler les fichiers de données dans un ou plusieurs volumes dédiés. Ils doivent être non contaminés par tout autre type de fichier. Cela permet de s'assurer que les volumes de fichiers de données peuvent être rapidement restaurés lors d'une opération SnapRestore sans détruire un journal de reprise, un fichier de contrôle ou un journal d'archivage important.

LE SYSTÈME SAN présente des exigences similaires en matière d'isolation des fichiers de données dans des volumes dédiés. Avec un système d'exploitation tel que Microsoft Windows, un seul volume peut contenir plusieurs LUN de fichiers de données, chacune avec un système de fichiers NTFS. Avec d'autres systèmes d'exploitation, il existe généralement un gestionnaire de volumes logiques. Par exemple, avec Oracle ASM, l'option la plus simple consiste à limiter les groupes de disques à un volume unique pouvant être sauvegardé et restauré comme une unité. Si des volumes supplémentaires sont nécessaires pour des raisons de performance ou de gestion de la capacité, la création d'un groupe de disques supplémentaire sur le nouveau volume simplifie la gestion.

Si ces instructions sont respectées, les snapshots peuvent être planifiés directement sur ONTAP sans avoir à créer de snapshot de groupe de cohérence. En effet, les sauvegardes optimisées pour les snapshots ne nécessitent pas la sauvegarde simultanée de fichiers de données.

Une complication se produit dans des situations telles qu'un groupe de disques ASM distribué sur des volumes. Dans ce cas, un snapshot de groupe de cohérence doit être réalisé pour s'assurer que les

métadonnées ASM sont cohérentes sur tous les volumes constitutifs.

[Remarque]Vérifiez que les fichiers spfile et passwd ASM ne se trouvent pas dans le groupe de disques hébergeant les fichiers de données. Cela interfère avec la capacité à restaurer de manière sélective les fichiers de données et uniquement les fichiers de données.

Procédure de restauration locale : NFS

Cette procédure peut être conduite manuellement ou via une application telle que SnapCenter. La procédure de base est la suivante :

1. Arrêtez la base de données.
2. Restaurez le ou les volumes de fichiers de données sur l'instantané immédiatement avant le point de restauration souhaité.
3. Réexécutez les journaux d'archivage au point souhaité.

Cette procédure suppose que les journaux d'archive souhaités sont toujours présents dans le système de fichiers actif. Si ce n'est pas le cas, les journaux d'archive doivent être restaurés, ou `rman` ou `sqlplus` peut être dirigé vers les données dans le `.snapshot` répertoire.

En outre, dans le cas de bases de données plus petites, l'utilisateur peut restaurer les fichiers de données directement à partir du système `.snapshot` Répertoire n'ayant pas besoin des outils d'automatisation ou d'un administrateur du stockage pour exécuter une commande SnapRestore.

Procédure de restauration locale—SAN

Cette procédure peut être conduite manuellement ou via une application telle que SnapCenter. La procédure de base est la suivante :

1. Arrêtez la base de données.
2. Arrêter le ou les groupes de disques hébergeant les fichiers de données. La procédure varie en fonction du gestionnaire de volumes logiques choisi. Avec ASM, le processus nécessite de démonter le groupe de disques. Sous Linux, les systèmes de fichiers doivent être démontés et les volumes logiques et les groupes de volumes désactivés. L'objectif est d'arrêter toutes les mises à jour du groupe de volumes cible à restaurer.
3. Restaurez les groupes de disques de fichiers de données sur l'instantané immédiatement avant le point de restauration souhaité.
4. Réactivez les groupes de disques récemment restaurés.
5. Réexécutez les journaux d'archivage au point souhaité.

Cette procédure suppose que les journaux d'archive souhaités sont toujours présents dans le système de fichiers actif. Si ce n'est pas le cas, les journaux d'archivage doivent être restaurés en mettant les LUN du journal d'archivage hors ligne et en effectuant une restauration. Il s'agit également d'un exemple dans lequel il est utile de diviser les journaux d'archivage en volumes dédiés. Si les journaux d'archivage partagent un groupe de volumes avec les journaux de reprise, les journaux de reprise doivent être copiés ailleurs avant la restauration de l'ensemble global de LUN afin d'éviter de perdre les transactions enregistrées finales.

Exemple de récupération complète

Supposons que les fichiers de données ont été corrompus ou détruits et qu'une restauration complète est requise. La procédure à suivre est la suivante :

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover automatic;
Media recovery complete.
SQL> alter database open;
Database altered.
SQL>
```

Exemple de restauration instantanée

Toute la procédure de restauration est une commande unique : `recover automatic`.

Si une restauration à un point dans le temps est requise, l'horodatage des snapshots doit être connu et peut être identifié comme suit :

```
Cluster01::> snapshot show -vserver vserver1 -volume NTAP_oradata -fields
create-time
vserver   volume        snapshot   create-time
--------  ------------  ---------  ------------------------
vserver1  NTAP_oradata  my-backup  Thu Mar 09 10:10:06 2017
```

L'heure de création de l'instantané est répertoriée comme 9 mars et 10:10:06. Pour être sûr, une minute est ajoutée à l'heure du snapshot :

```
[oracle@host1 ~]$ sqlplus / as sysdba
Connected to an idle instance.
SQL> startup mount;
ORACLE instance started.
Total System Global Area 1610612736 bytes
Fixed Size                  2924928 bytes
Variable Size            1040191104 bytes
Database Buffers          553648128 bytes
Redo Buffers               13848576 bytes
Database mounted.
SQL> recover database until time '09-MAR-2017 10:44:15' snapshot time '09-
MAR-2017 10:11:00';
```

La restauration est maintenant lancée. Il a spécifié une heure d'instantané de 10:11:00, une minute après l'heure enregistrée pour tenir compte de la variation d'horloge possible, et un temps de récupération cible de 10:44. Ensuite, sqlplus demande les journaux d'archivage requis pour atteindre le délai de restauration souhaité de 10:44.

```
ORA-00279: change 551760 generated at 03/09/2017 05:06:07 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_31_930813377.dbf
ORA-00280: change 551760 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 552566 generated at 03/09/2017 05:08:09 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_32_930813377.dbf
ORA-00280: change 552566 for thread 1 is in sequence #32
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 553045 generated at 03/09/2017 05:10:12 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_33_930813377.dbf
ORA-00280: change 553045 for thread 1 is in sequence #33
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 753229 generated at 03/09/2017 05:15:58 needed for
thread 1
ORA-00289: suggestion : /oralogs_nfs/arch/1_34_930813377.dbf
ORA-00280: change 753229 for thread 1 is in sequence #34
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
Log applied.
Media recovery complete.
SQL> alter database open resetlogs;
Database altered.
SQL>
```

Restauration complète d'une base de données à l'aide de snapshots à l'aide de `recover automatic` la commande ne nécessite pas de licence spécifique, mais une restauration à un point dans le temps via `snapshot time` Requiert la licence Oracle Advanced compression.

Outils d'automatisation et de gestion de la base de données

Dans un environnement de base de données Oracle, la principale valeur de ONTAP provient des principales technologies ONTAP, telles que les copies Snapshot instantanées, la réplication simple SnapMirror et la création efficace de volumes FlexClone.

Dans certains cas, une configuration simple de ces fonctionnalités principales directement sur ONTAP répond aux exigences, mais les besoins plus complexes requièrent une couche d'orchestration.

SnapCenter

SnapCenter est le produit phare de la protection des données NetApp. À un niveau très bas, il est similaire aux produits SnapManager en termes d'exécution des sauvegardes de base de données, mais il a été conçu dès le départ pour proposer une gestion de la protection des données centralisée sur les systèmes de stockage NetApp.

SnapCenter inclut les fonctions de base telles que les sauvegardes et restaurations basées sur des snapshots, SnapMirror et la réplication SnapVault, ainsi que d'autres fonctionnalités nécessaires pour fonctionner à grande échelle pour les grandes entreprises. Ces fonctionnalités avancées incluent un contrôle d'accès basé sur des rôles (RBAC) étendu, des API RESTful pour l'intégration de produits d'orchestration tiers, une gestion centralisée et sans interruption des plug-ins SnapCenter sur des hôtes de base de données et une interface utilisateur conçue pour les environnements à l'échelle du cloud.

REPOS

ONTAP contient également un jeu d'API RESTful riche. Les fournisseurs tiers peuvent ainsi créer une application de protection des données et de gestion grâce à une intégration étroite avec ONTAP. De plus, l'API RESTful est facile à utiliser par les clients qui souhaitent créer leurs propres workflows et utilitaires d'automatisation.

Reprise sur incident Oracle

Présentation

La reprise d'activité consiste à restaurer les services de données après une catastrophe, par exemple un incendie qui détruit un système de stockage, voire un site entier.

> Cette documentation remplace les rapports techniques *TR-4591 : Oracle Data protection* et *TR-4592 : Oracle on MetroCluster.*

La reprise après incident peut être effectuée par une simple réplication des données à l'aide de SnapMirror, bien sûr, lorsque de nombreux clients mettent à jour les réplicas en miroir toutes les heures.

Pour la plupart des clients, la reprise après incident ne suffit pas à posséder une copie distante des données. Il est donc nécessaire de pouvoir les exploiter rapidement. NetApp propose deux technologies pour répondre à ce besoin : MetroCluster et SnapMirror Active Sync

MetroCluster fait référence à ONTAP dans une configuration matérielle qui inclut un stockage en miroir synchrone de faible niveau et de nombreuses fonctionnalités supplémentaires. Les solutions intégrées telles que MetroCluster simplifient les bases de données, les applications et les infrastructures de virtualisation complexes et évolutives. Elle remplace plusieurs produits et stratégies externes de protection des données par une seule baie de stockage centrale simple. Elle offre également des fonctionnalités intégrées de sauvegarde, de restauration, de reprise après incident et de haute disponibilité au sein d'un seul système de stockage en cluster.

La synchronisation active SnapMirror (SM-AS) est basée sur la synchronisation SnapMirror synchrone. Avec MetroCluster, chaque contrôleur ONTAP est responsable de la réplication des données de son disque vers un emplacement distant. Avec la synchronisation active SnapMirror, deux systèmes ONTAP différents conservent des copies indépendantes de vos données LUN, mais fonctionnent ensemble pour présenter une seule instance de ce LUN. Du point de vue de l'hôte, il s'agit d'une entité LUN unique.

Comparaison SM-AS et MCC

Si les solutions SM-AS et MetroCluster sont similaires en termes de fonctionnalité globale, elles présentent d'importantes différences dans la mise en œuvre de la réplication avec un objectif de point de récupération de 0 et sa gestion. Les modes asynchrone et synchrone de SnapMirror peuvent également être utilisés dans le cadre d'un plan de reprise d'activité, mais ils ne sont pas conçus pour être utilisés en tant que technologies de réplication haute disponibilité.

- Une configuration MetroCluster ressemble davantage à un cluster intégré avec des nœuds distribués sur plusieurs sites. SM-AS se comporte comme deux clusters indépendants qui coopèrent pour fournir des LUN répliquées synchrones avec RPO=0 sélectionnés.
- Les données d'une configuration MetroCluster ne sont accessibles qu'à partir d'un site particulier à la fois. Une deuxième copie des données est présente sur le site opposé, mais les données sont passives. Il est impossible d'y accéder sans un basculement du système de stockage.
- La mise en miroir des systèmes MetroCluster et SM-AS effectue des opérations à différents niveaux. La mise en miroir MetroCluster s'effectue au niveau de la couche RAID. Les données de bas niveau sont stockées dans un format miroir à l'aide de SyncMirror. L'utilisation de la mise en miroir est pratiquement invisible au niveau des couches LUN, volume et protocole.
- En revanche, la mise en miroir SM-AS se produit au niveau de la couche de protocole. Les deux clusters sont globalement indépendants. Une fois les deux copies de données synchronisées, les deux clusters n'ont besoin que de mettre en miroir les écritures. Lorsqu'une écriture a lieu sur un cluster, elle est répliquée sur l'autre. L'écriture est uniquement validée par l'hôte lorsque l'écriture est terminée sur les deux sites. En dehors de ce comportement de fractionnement de protocole, les deux clusters sont des clusters ONTAP normaux.
- Le rôle principal de MetroCluster est la réplication à grande échelle. Vous pouvez répliquer une baie complète avec un objectif de point de récupération RPO=0 et un objectif de durée de restauration proche de zéro. Le processus de basculement est ainsi simplifié, car il n'y a qu'une seule « chose » à basculer et il offre une excellente évolutivité en termes de capacité et d'IOPS.
- L'une des principales utilisations de SM-AS est la réplication granulaire. Parfois, vous ne souhaitez pas répliquer toutes les données en tant qu'unité unique ou vous devez pouvoir basculer sélectivement sur certains workloads.
- Autre cas d'utilisation clé de la solution SM-as pour les opérations actives/actives : vous souhaitez que des copies de données entièrement exploitables soient disponibles sur deux clusters différents situés à deux emplacements différents avec des performances identiques et, si vous le souhaitez, vous n'avez pas besoin d'étendre le SAN sur plusieurs sites. Vos applications peuvent déjà s'exécuter sur les deux sites, ce qui réduit le RTO global pendant les opérations de basculement.

MetroCluster

Reprise d'activité avec MetroCluster

MetroCluster est une fonctionnalité ONTAP qui protège vos bases de données Oracle avec une mise en miroir synchrone RPO=0 sur tous les sites. Elle peut évoluer jusqu'à prendre en charge des centaines de bases de données sur un seul système MetroCluster.

Il est également simple à utiliser. L'utilisation de MetroCluster n'ajoute pas nécessairement à ou ne modifie pas nécessairement les meilleures ractiques pour l'exploitation des applications et bases de données d'entreprise.

Les bonnes pratiques habituelles s'appliquent toujours. Si vos besoins requièrent uniquement une protection des données avec un objectif de point de récupération de 0, MetroCluster répond à ce besoin. Cependant, la

plupart des clients utilisent MetroCluster non seulement pour la protection des données avec un objectif de point de récupération de 0, mais aussi pour améliorer l'objectif de délai de restauration en cas d'incident et fournir un basculement transparent dans le cadre des activités de maintenance du site.

Architecture physique

Pour comprendre le fonctionnement des bases de données Oracle dans un environnement MetroCluster, il est nécessaire d'expliquer la conception physique d'un système MetroCluster.

> Cette documentation remplace le rapport technique *TR-4592 : Oracle on MetroCluster.*

MetroCluster est disponible dans 3 configurations différentes

- Paires HAUTE DISPONIBILITÉ avec connectivité IP
- Paires HAUTE DISPONIBILITÉ avec connectivité FC
- Contrôleur unique avec connectivité FC

> Le terme « connectivité » fait référence à la connexion au cluster utilisée pour la réplication entre sites. Il ne fait pas référence aux protocoles hôtes. Tous les protocoles côté hôte sont pris en charge comme d'habitude dans une configuration MetroCluster, quel que soit le type de connexion utilisé pour les communications entre clusters.

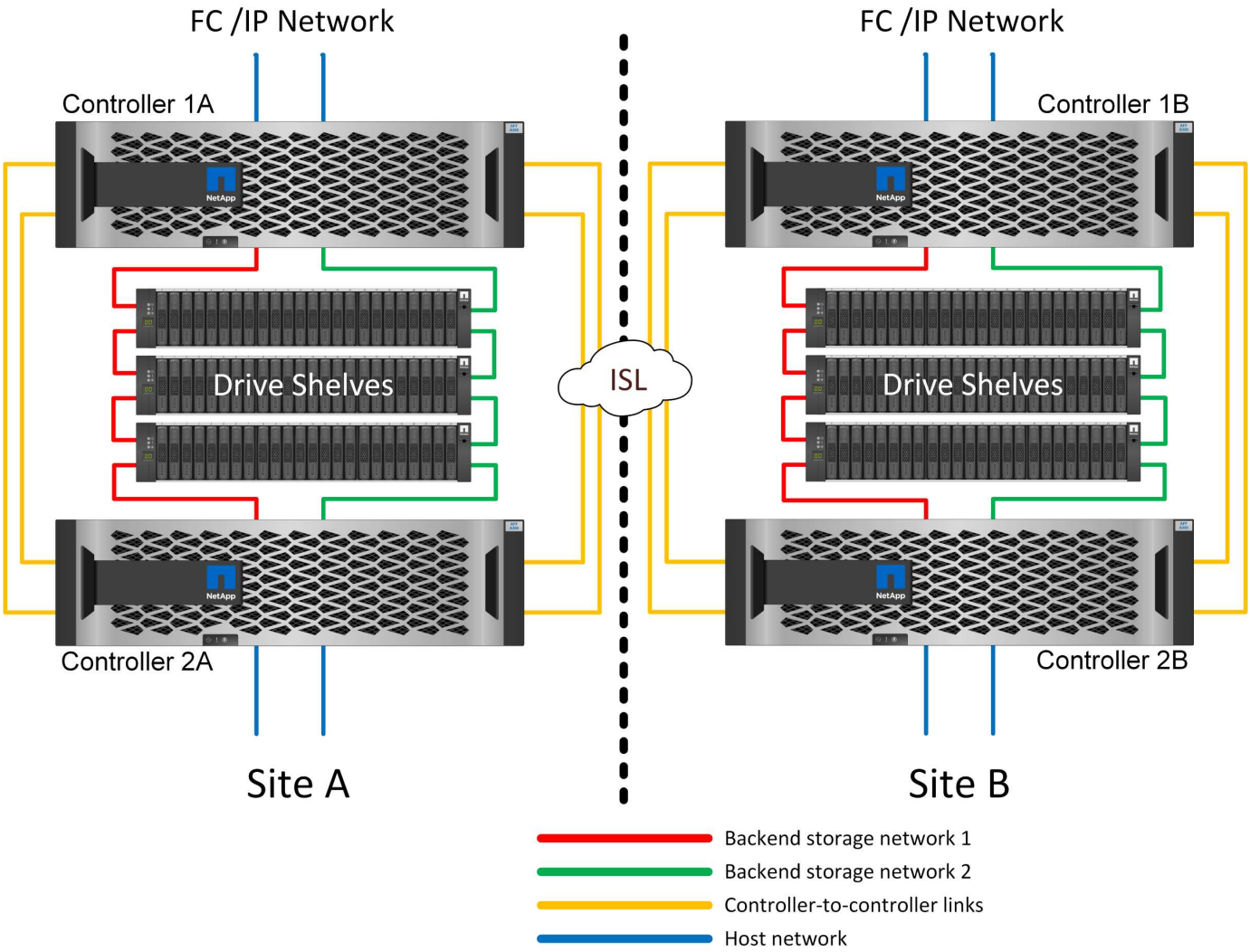
IP MetroCluster

La configuration IP MetroCluster à paire haute disponibilité utilise deux ou quatre nœuds par site. Cette option de configuration augmente la complexité et les coûts liés à l'option à deux nœuds, mais elle offre un avantage important : la redondance intrasite. Une simple panne de contrôleur ne nécessite pas l'accès aux données via le WAN. L'accès aux données reste local via l'autre contrôleur local.

La plupart des clients choisissent la connectivité IP, car les exigences d'infrastructure sont plus simples. Auparavant, la connectivité inter-sites à haut débit était généralement plus facile à provisionner avec des commutateurs FC et fibre noire. Cependant, les circuits IP à haut débit et à faible latence sont aujourd'hui plus facilement disponibles.

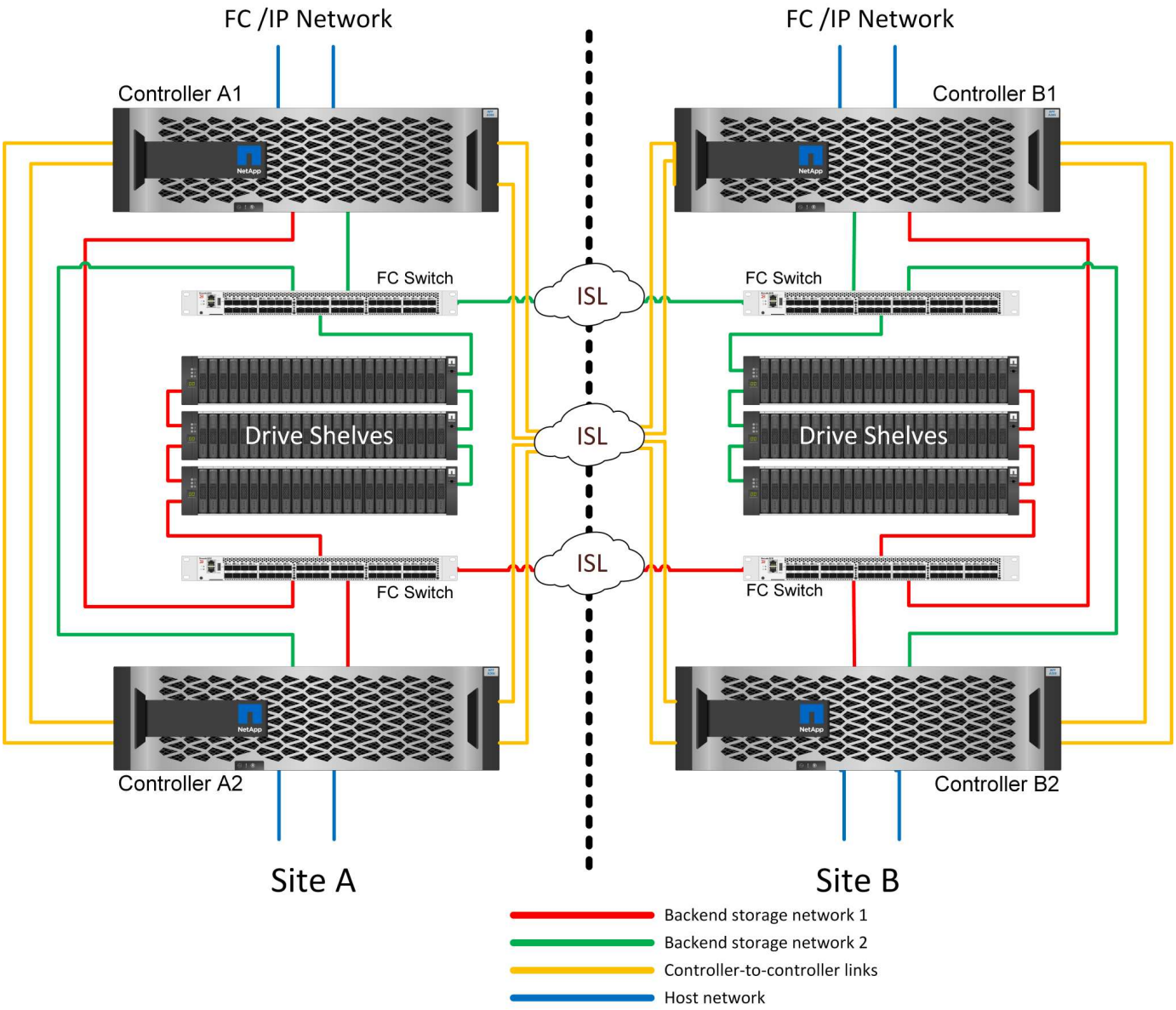
L'architecture est également plus simple, car les contrôleurs disposent des seules connexions entre les sites. Dans les MetroCluster FC, un contrôleur écrit directement sur les disques du site opposé et requiert ainsi des connexions SAN, des commutateurs et des ponts supplémentaires. En revanche, un contrôleur dans une configuration IP écrit sur les lecteurs opposés via le contrôleur.

Pour plus d'informations, consultez la documentation officielle de ONTAP et "Architecture et conception de la solution IP de MetroCluster".



MetroCluster FC à connexion SAN HA-pair

La configuration MetroCluster FC à paire haute disponibilité utilise deux ou quatre nœuds par site. Cette option de configuration augmente la complexité et les coûts liés à l'option à deux nœuds, mais elle offre un avantage important : la redondance intrasite. Une simple panne de contrôleur ne nécessite pas l'accès aux données via le WAN. L'accès aux données reste local via l'autre contrôleur local.



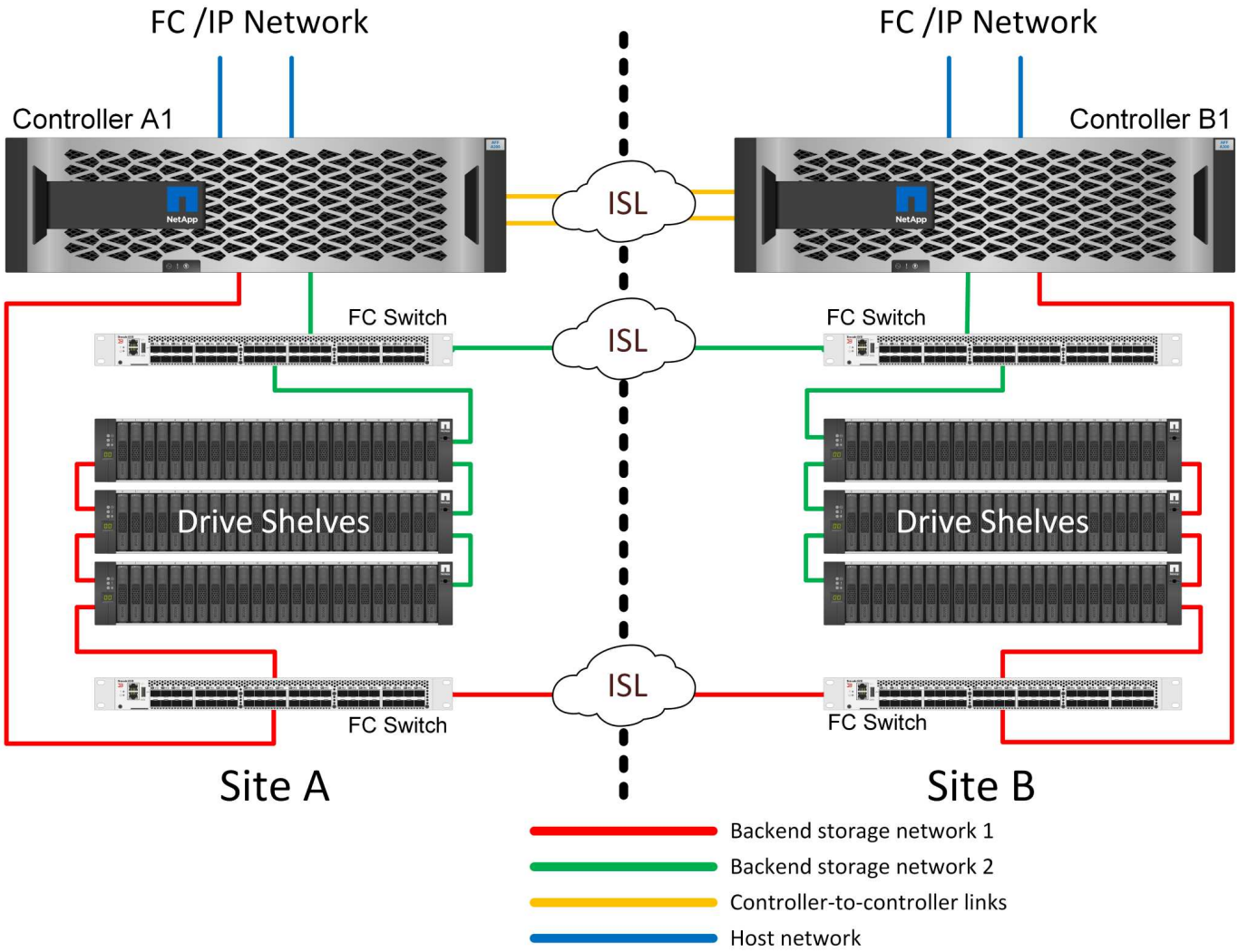
Certaines infrastructures multisites ne sont pas conçues pour les opérations en mode actif-actif. Elles sont plutôt utilisées comme site principal et site de reprise après incident. Dans ce cas, il est généralement préférable d'utiliser une option MetroCluster à paire HA pour les raisons suivantes :

- Bien qu'un cluster MetroCluster à deux nœuds soit un système haute disponibilité, toute panne inattendue d'un contrôleur ou une maintenance planifiée implique que les services de données soient en ligne sur le site opposé. Si la connectivité réseau entre les sites ne prend pas en charge la bande passante requise, les performances sont affectées. La seule option serait également de basculer les différents systèmes d'exploitation hôtes et les services associés vers le site secondaire. Le cluster MetroCluster de paire haute disponibilité élimine ce problème, car la perte d'un contrôleur simplifie le basculement au sein du même site.
- Certaines topologies réseau ne sont pas conçues pour l'accès intersite, mais utilisent des sous-réseaux différents ou des SAN FC isolés. Dans ce cas, le cluster MetroCluster à deux nœuds ne fonctionne plus comme un système haute disponibilité, car le contrôleur secondaire ne peut plus transmettre de données aux serveurs sur le site opposé. L'option MetroCluster de paire haute disponibilité est nécessaire pour assurer une redondance complète.
- Si une infrastructure à deux sites est considérée comme une seule infrastructure extrêmement disponible, la configuration MetroCluster à deux nœuds est adaptée. Toutefois, si le système doit fonctionner pendant

une période prolongée après une panne sur le site, une paire haute disponibilité est recommandée, car la haute disponibilité continue d'être disponible sur un seul site.

MetroCluster FC à deux nœuds avec connexion SAN

La configuration MetroCluster à deux nœuds n'utilise qu'un nœud par site. Cette conception est plus simple que l'option de paire haute disponibilité, car le nombre de composants à configurer et à gérer est inférieur. Elle a également réduit les besoins en infrastructure en termes de câblage et de commutation FC. Enfin, il réduit les coûts.



L'impact évident de cette conception est que la défaillance du contrôleur sur un seul site signifie que les données sont disponibles depuis le site opposé. Cette restriction n'est pas nécessairement un problème. De nombreuses entreprises disposent d'opérations de data Center multisites avec des réseaux étendus, ultra-rapides et à faible latence qui fonctionnent essentiellement comme une infrastructure unique. Dans ce cas, la version à deux nœuds de MetroCluster est la configuration préférée. Plusieurs fournisseurs de services utilisent actuellement des systèmes à deux nœuds de plusieurs pétaoctets.

Fonctions de résilience MetroCluster

Une solution MetroCluster ne présente aucun point de défaillance unique :

- Chaque contrôleur dispose de deux chemins d'accès indépendants aux tiroirs disques sur le site local.

- Chaque contrôleur dispose de deux chemins d'accès indépendants aux tiroirs disques du site distant.
- Chaque contrôleur dispose de deux chemins d'accès indépendants aux contrôleurs sur le site opposé.
- Dans la configuration HA-pair, chaque contrôleur dispose de deux chemins vers son partenaire local.

En résumé, n'importe quel composant de la configuration peut être supprimé sans compromettre la capacité de MetroCluster à transmettre des données. La seule différence en termes de résilience entre les deux options est que la version à paire haute disponibilité reste un système de stockage haute disponibilité global après une panne de site.

Architecture logique

Comprendre le fonctionnement des bases de données Oracle dans un environnement MetroCluster alsop nécessite une explication de la fonctionnalité logique d'un système MetroCluster.

Protection contre les défaillances de site : NVRAM et MetroCluster

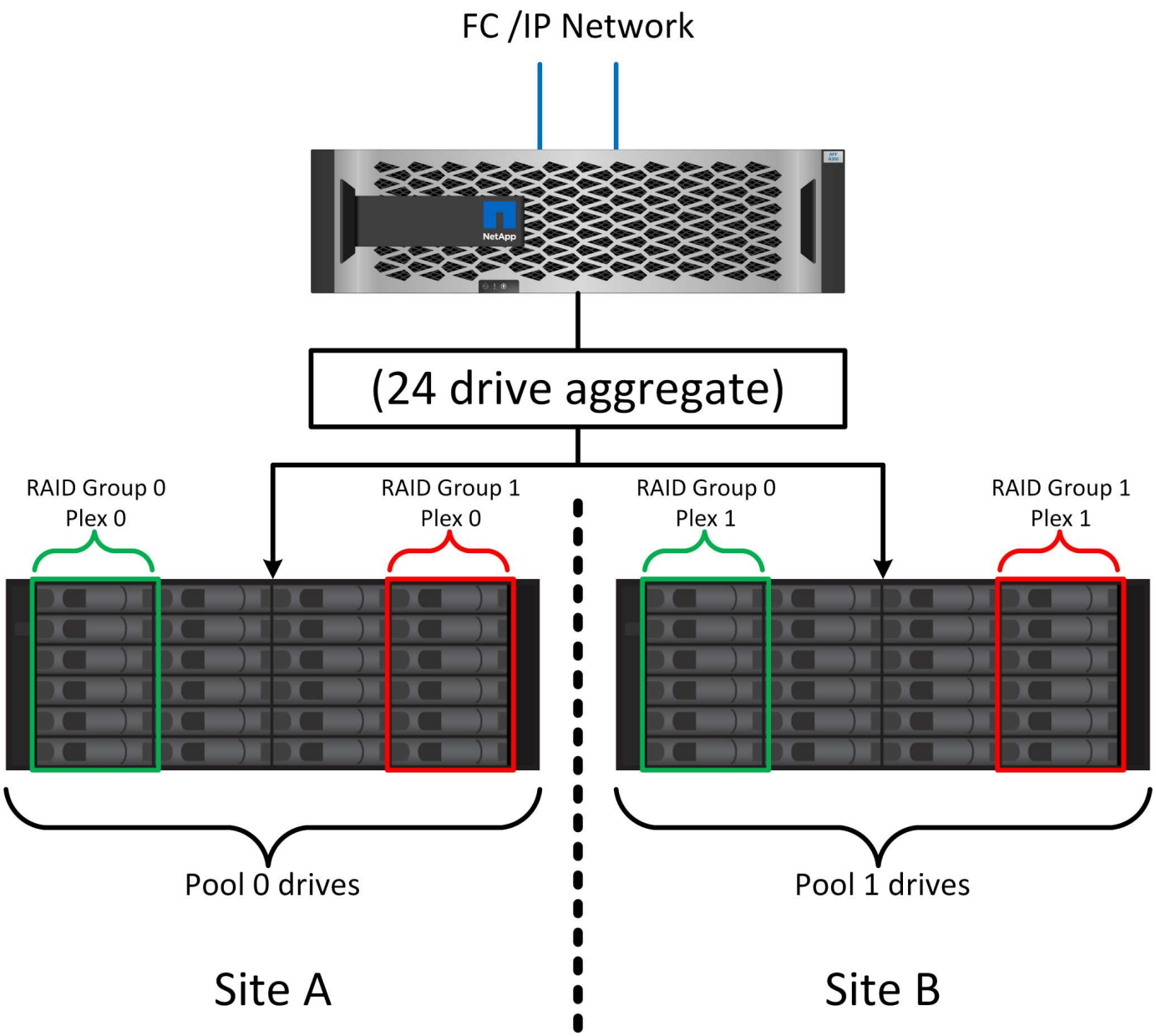
MetroCluster étend la protection des données NVRAM de plusieurs manières :

- Dans une configuration à deux nœuds, les données NVRAM sont répliquées au partenaire distant à l'aide des liens ISL (Inter-Switch Links).
- Dans une configuration de paire haute disponibilité, les données NVRAM sont répliquées à la fois vers le partenaire local et vers un partenaire distant.
- Une écriture n'est pas validée tant qu'elle n'est pas répliquée à tous les partenaires. Cette architecture protège les E/S à la volée contre les défaillances de site en répliquant les données NVRAM sur un partenaire distant. Ce processus n'est pas impliqué dans la réplication des données au niveau des disques. Le contrôleur propriétaire des agrégats est responsable de la réplication des données en écrivant dans les deux plexes de l'agrégat. Cependant, il doit toujours assurer une protection contre les pertes d'E/S à la volée en cas de perte du site. Les données NVRAM répliquées sont uniquement utilisées si un contrôleur partenaire doit prendre le relais en cas de défaillance d'un contrôleur.

Protection contre les pannes de site et de tiroir : SyncMirror et plexes

SyncMirror est une technologie de mise en miroir qui améliore, mais ne remplace pas, RAID DP ou RAID-TEC. Il met en miroir le contenu de deux groupes RAID indépendants. La configuration logique est la suivante :

1. Les disques sont configurés en deux pools en fonction de leur emplacement. Un pool est composé de tous les disques du site A et le second est composé de tous les disques du site B.
2. Un pool de stockage commun, appelé agrégat, est ensuite créé à partir de jeux en miroir de groupes RAID. Un nombre égal de lecteurs est tiré de chaque site. Par exemple, un agrégat SyncMirror de 20 disques se compose de 10 disques du site A et de 10 disques du site B.
3. Chaque jeu de disques d'un site donné est automatiquement configuré comme un ou plusieurs groupes RAID DP ou RAID-TEC entièrement redondants, indépendamment de l'utilisation de la mise en miroir. Cette utilisation de la mise en miroir RAID assure la protection des données même après la perte d'un site.



La figure ci-dessus illustre un exemple de configuration SyncMirror. Un agrégat de 24 disques a été créé sur le contrôleur avec 12 disques à partir d'un tiroir alloué sur le site A et 12 disques à partir d'un tiroir alloué sur le site B. Les disques ont été regroupés en deux groupes RAID en miroir. Le groupe RAID 0 comprend un plex de 6 disques sur le site A mis en miroir sur un plex de 6 disques sur le site B. De même, le groupe RAID 1 comprend un plex de 6 disques sur le site A mis en miroir sur un plex de 6 disques sur le site B.

SyncMirror est généralement utilisé pour assurer la mise en miroir à distance avec les systèmes MetroCluster, avec une copie des données sur chaque site. Il a parfois été utilisé pour fournir un niveau supplémentaire de redondance dans un seul système. Il assure en particulier la redondance au niveau du tiroir. Un tiroir disque contient déjà deux blocs d'alimentation et contrôleurs. Dans l'ensemble, il ne s'agit pas d'une simple tôlerie, mais dans certains cas, une protection supplémentaire peut être garantie. Par exemple, un client NetApp a déployé SyncMirror sur une plateforme mobile d'analytique en temps réel utilisée lors des tests automobiles. Le système a été séparé en deux racks physiques fournis avec des alimentations indépendantes et des systèmes UPS indépendants.

Échec de la redondance : NVFAIL

Comme nous l'avons vu précédemment, une écriture n'est pas validée tant qu'elle n'a pas été connectée à la NVRAM et à la NVRAM locales sur au moins un autre contrôleur. Cette approche évite toute panne matérielle ou de courant qui entraîne une perte des E/S à la volée En cas de panne de la mémoire NVRAM locale ou de la connectivité aux autres nœuds, les données ne seront plus mises en miroir.

Si la mémoire NVRAM locale signale une erreur, le nœud s'arrête. Cet arrêt entraîne le basculement vers un contrôleur partenaire lorsque des paires haute disponibilité sont utilisées. Avec MetroCluster, le comportement dépend de la configuration globale choisie, mais il peut entraîner un basculement automatique vers la note distante. Dans tous les cas, aucune donnée n'est perdue parce que le contrôleur qui connaît la défaillance n'a pas acquitté l'opération d'écriture.

Une défaillance de connectivité site à site qui bloque la réplication NVRAM sur des nœuds distants est une situation plus compliquée. Les écritures ne sont plus répliquées sur les nœuds distants, ce qui crée un risque de perte de données en cas d'erreur catastrophique sur un contrôleur. Plus important encore, une tentative de basculement vers un autre nœud dans ces conditions entraîne une perte de données.

Le facteur de contrôle est de savoir si la NVRAM est synchronisée. Si la mémoire NVRAM est synchronisée, le basculement nœud à nœud peut se poursuivre sans risque de perte de données. Dans une configuration MetroCluster, si la mémoire NVRAM et les plexes d'agrégats sous-jacents sont synchronisés, vous pouvez procéder au basculement sans risque de perte de données.

ONTAP n'autorise pas le basculement ou le basculement lorsque les données ne sont pas synchronisées, sauf si le basculement ou le basculement est forcé. Le fait de forcer une modification des conditions de cette manière reconnaît que les données peuvent être laissées pour compte dans le contrôleur d'origine et que la perte de données est acceptable.

Les bases de données et autres applications sont particulièrement vulnérables à la corruption en cas de basculement ou de basculement forcé, car elles conservent des caches internes de données plus volumineux sur disque. En cas de basculement forcé ou de basculement forcé, les modifications précédemment reconnues sont effectivement supprimées. Le contenu de la baie de stockage recule dans le temps et l'état du cache ne reflète plus l'état des données sur le disque.

Afin d'éviter ce genre de situation, ONTAP permet de configurer les volumes pour une protection spéciale contre les défaillances de mémoire NVRAM. Lorsqu'il est déclenché, ce mécanisme de protection entraîne l'entrée d'un volume dans un état appelé NVFAIL. Cet état entraîne des erreurs d'E/S qui provoquent une panne de l'application. Cette panne provoque l'arrêt des applications, qui n'utilisent donc pas de données obsolètes. Les données ne doivent pas être perdues car des données de transaction validées doivent être présentes dans les journaux. Les étapes suivantes habituelles sont qu'un administrateur arrête complètement les hôtes avant de remettre manuellement en ligne les LUN et les volumes. Bien que ces étapes puissent impliquer un certain travail, cette approche est le moyen le plus sûr d'assurer l'intégrité des données. Toutes les données n'ont pas besoin de cette protection. C'est pourquoi NVFAIL peut être configuré volume par volume.

Paires HAUTE DISPONIBILITÉ et MetroCluster

MetroCluster est disponible dans deux configurations : deux nœuds et paire haute disponibilité. La configuration à deux nœuds se comporte de la même manière qu'une paire haute disponibilité par rapport à la mémoire NVRAM. En cas de défaillance soudaine, le nœud partenaire peut relire les données NVRAM pour assurer la cohérence des disques et garantir la perte d'aucune écriture reconnue.

La configuration HA-pair réplique également la mémoire NVRAM sur le nœud partenaire local. Une simple défaillance de contrôleur entraîne une relecture NVRAM sur le nœud partenaire, comme c'est le cas avec une paire haute disponibilité autonome sans MetroCluster. En cas de perte complète soudaine d'un site, le site

distant dispose également de la mémoire NVRAM requise pour assurer la cohérence des disques et commencer à transmettre les données.

Un aspect important de MetroCluster est que les nœuds distants ne peuvent pas accéder aux données des partenaires dans des conditions de fonctionnement normales. Chaque site fonctionne essentiellement comme un système indépendant qui peut assumer la personnalité du site opposé. Ce processus est connu sous le nom de basculement et inclut un basculement planifié dans lequel les opérations sur site sont migrées sans interruption vers le site opposé. Il comprend également les situations non planifiées où un site est perdu et un basculement manuel ou automatique est nécessaire dans le cadre de la reprise d'activité.

Basculement et rétablissement

Les termes « switchover and switchoback » font référence au processus de transition des volumes entre des contrôleurs distants dans une configuration MetroCluster. Ce processus s'applique uniquement aux nœuds distants. Lorsque MetroCluster est utilisé dans une configuration à quatre volumes, le basculement de nœud local est le même processus de basculement et de rétablissement que celui décrit précédemment.

Basculement et rétablissement planifiés

Un basculement ou rétablissement planifié est similaire à un basculement ou un rétablissement entre les nœuds. Ce processus comporte plusieurs étapes et peut sembler prendre plusieurs minutes, mais il s'agit d'une transition progressive et progressive des ressources de stockage et de réseau. Le moment où les transferts de contrôle se produisent beaucoup plus rapidement que le temps nécessaire à l'exécution de la commande complète.

La principale différence entre le basculement/rétablissement et le basculement/rétablissement réside dans l'effet sur la connectivité FC SAN. Avec le Takeover/Giveback local, un hôte subit la perte de tous les chemins FC vers le nœud local et s'appuie sur son MPIO natif pour le basculer vers des chemins alternatifs disponibles. Les ports ne sont pas déplacés. Avec le basculement et le rétablissement, les ports cibles FC virtuels des contrôleurs passent à l'autre site. Ils cessent d'exister sur le SAN pendant un instant, puis réapparaissent sur un autre contrôleur.

SyncMirror expire

SyncMirror est une technologie de mise en miroir ONTAP qui offre une protection contre les défaillances de tiroirs. Lorsque les tiroirs sont séparés sur une distance, les données sont protégées à distance.

SyncMirror ne fournit pas de mise en miroir synchrone universelle. Le résultat est une meilleure disponibilité. Certains systèmes de stockage utilisent une mise en miroir totale ou nulle constante, parfois appelée mode domino. Cette forme de mise en miroir est limitée dans l'application car toutes les activités d'écriture doivent cesser en cas de perte de la connexion au site distant. Sinon, une écriture existerait sur un site, mais pas sur l'autre. Généralement, ces environnements sont configurés pour mettre les LUN hors ligne en cas de perte de la connectivité site à site pendant plus d'une courte période (par exemple, 30 secondes).

Ce comportement est souhaitable pour un petit sous-ensemble d'environnements. Cependant, la plupart des applications nécessitent une solution capable de garantir une réplication synchrone dans des conditions normales de fonctionnement, mais avec la possibilité de suspendre la réplication. Une perte complète de la connectivité site à site est souvent considérée comme une situation proche d'une catastrophe. Généralement, ces environnements sont maintenus en ligne et donnent accès aux données jusqu'à ce que la connectivité soit réparée ou qu'une décision officielle soit prise de fermer l'environnement pour protéger les données. Il n'est pas rare d'avoir besoin d'arrêter automatiquement l'application uniquement en raison d'une défaillance de réplication à distance.

SyncMirror prend en charge les exigences de mise en miroir synchrone avec la flexibilité d'un délai d'expiration. Si la connectivité à la télécommande et/ou au plex est perdue, une minuterie de 30 secondes

commence à s'arrêter. Lorsque le compteur atteint 0, le traitement des E/S d'écriture reprend en utilisant les données locales. La copie distante des données est utilisable, mais elle est figée à temps jusqu'à ce que la connectivité soit rétablie. La resynchronisation exploite des snapshots au niveau de l'agrégat pour rétablir le système en mode synchrone aussi rapidement que possible.

Notamment, dans de nombreux cas, ce type de réplication universelle en mode domino tout ou rien est mieux implémenté au niveau de la couche applicative. Par exemple, Oracle DataGuard inclut le mode de protection maximum, ce qui garantit la réplication à long terme en toutes circonstances. Si la liaison de réplication échoue pendant une période dépassant un délai configurable, les bases de données s'arrêtent.

Basculement automatique sans surveillance avec Fabric Attached MetroCluster

Le basculement automatique sans surveillance (AUSO) est une fonctionnalité MetroCluster intégrée au fabric qui offre une forme de haute disponibilité intersite. Comme évoqué précédemment, MetroCluster est disponible en deux types : un contrôleur unique sur chaque site ou une paire haute disponibilité sur chaque site. L'avantage principal de l'option haute disponibilité est que l'arrêt planifié ou non planifié du contrôleur permet toujours une E/S locale. L'avantage de l'option à nœud unique est de réduire les coûts, la complexité et l'infrastructure.

La principale valeur d'AUSO est d'améliorer les fonctionnalités haute disponibilité des systèmes MetroCluster connectés à la structure. Chaque site surveille l'état de santé du site opposé et, si aucun nœud n'est encore utilisé pour transmettre des données, l'AUSO assure un basculement rapide. Cette approche est particulièrement utile dans les configurations MetroCluster avec un seul nœud par site, car elle rapproche la configuration d'une paire haute disponibilité en termes de disponibilité.

AUSO ne peut pas offrir de surveillance complète au niveau d'une paire HA. Une paire haute disponibilité peut offrir une haute disponibilité, car elle inclut deux câbles physiques redondants pour une communication nœud à nœud directe. En outre, les deux nœuds d'une paire haute disponibilité ont accès au même ensemble de disques sur des boucles redondantes, ce qui permet à un nœud de suivre l'état d'un autre nœud sur une autre route.

Il existe des clusters MetroCluster sur plusieurs sites pour lesquels la communication nœud à nœud et l'accès au disque reposent sur la connectivité réseau site à site. La capacité à surveiller le pouls du reste du cluster est limitée. AUSO doit faire la distinction entre une situation où l'autre site est en fait hors service plutôt qu'indisponible en raison d'un problème de réseau.

Par conséquent, un contrôleur d'une paire haute disponibilité peut demander un basculement s'il détecte une panne de contrôleur qui s'est produite pour une raison spécifique, par exemple une situation critique du système. Elle peut également déclencher un basculement en cas de perte complète de la connectivité, parfois appelée « perte de pulsation ».

Un système MetroCluster ne peut effectuer un basculement automatique en toute sécurité que lorsqu'une panne spécifique est détectée sur le site d'origine. En outre, le contrôleur qui devient propriétaire du système de stockage doit être en mesure de garantir la synchronisation des données du disque et de la NVRAM. Le contrôleur ne peut pas garantir la sécurité d'un basculement simplement parce qu'il a perdu le contact avec le site source, qui pourrait toujours être opérationnel. Pour plus d'informations sur les options d'automatisation d'un basculement, reportez-vous aux informations sur la solution MetroCluster Tiebreaker (MCTB) dans la section suivante.

Disjoncteur d'attache MetroCluster avec MetroCluster FAS

"NetApp MetroCluster Tiebreaker"Exécuté sur un troisième site, le logiciel peut contrôler l'état de santé de l'environnement MetroCluster, envoyer des notifications et forcer un basculement en cas d'incident. Une description complète du Tiebreaker se trouve sur le "Site de support NetApp", mais le but principal du Tiebreaker MetroCluster est de détecter la perte du site. Il doit également faire la distinction entre la perte du

site et une perte de connectivité. Par exemple, le basculement ne doit pas se produire car le disjoncteur d'attache n'a pas pu atteindre le site principal. C'est pourquoi le disjoncteur d'attache surveille également la capacité du site distant à contacter le site principal.

Le basculement automatique avec AUSO est également compatible avec le MCTB. AUSO réagit très rapidement car il est conçu pour détecter des événements de défaillance spécifiques, puis n'invoque le basculement que lorsque les plexes NVRAM et SyncMirror sont synchronisés.

En revanche, le disjoncteur principal est situé à distance et doit donc attendre qu'une minuterie s'écoule avant de déclarer un site mort. Le disjoncteur d'attache détecte finalement le type de défaillance de contrôleur couverte par l'AUSO, mais en général, l'AUSO a déjà commencé le basculement et éventuellement terminé le basculement avant que le disjoncteur d'attache n'agisse. La deuxième commande de basculement qui en résulte provient du Tiebreaker serait rejetée.

> Le logiciel MCTB ne vérifie pas que NVRAM était et/ou que les plexes sont synchronisés lorsqu'un basculement est forcé. Le basculement automatique, s'il est configuré, doit être désactivé pendant les opérations de maintenance qui entraînent une perte de synchronisation des plexes NVRAM ou SyncMirror.

En outre, le MCTB peut ne pas traiter un désastre roulant qui conduit à la séquence d'événements suivante :

1. La connectivité entre les sites est interrompue pendant plus de 30 secondes.
2. La réplication SyncMirror est obsolète et les opérations se poursuivent sur le site principal, ce qui ne permet pas au réplica distant d'être obsolète.
3. Le site primaire est perdu. Le résultat est la présence de modifications non répliquées sur le site primaire. Un basculement peut alors se révéler indésirable pour plusieurs raisons, notamment :
   - Certaines données critiques peuvent être présentes sur le site primaire et peuvent être récupérées à terme. Un basculement qui a permis à l'application de continuer à fonctionner aurait pour effet de supprimer ces données stratégiques.
   - Des données peuvent être mises en cache pour une application sur le site survivant qui utilisait des ressources de stockage sur le site principal au moment de la perte du site. Le basculement introduit une version obsolète des données qui ne correspond pas au cache.
   - Des données peuvent être mises en cache sur un système d'exploitation du site survivant qui utilisait des ressources de stockage sur le site principal au moment de la perte du site. Le basculement introduit une version obsolète des données qui ne correspond pas au cache. L'option la plus sûre est de configurer le Tiebreaker pour envoyer une alerte s'il détecte une défaillance du site et demander à une personne de décider si elle doit forcer un basculement. Il peut être nécessaire d'abord d'arrêter les applications et/ou les systèmes d'exploitation pour effacer les données en cache. En outre, les paramètres NVFAIL peuvent être utilisés pour renforcer la protection et rationaliser le processus de basculement.

Mediator ONTAP avec MetroCluster IP

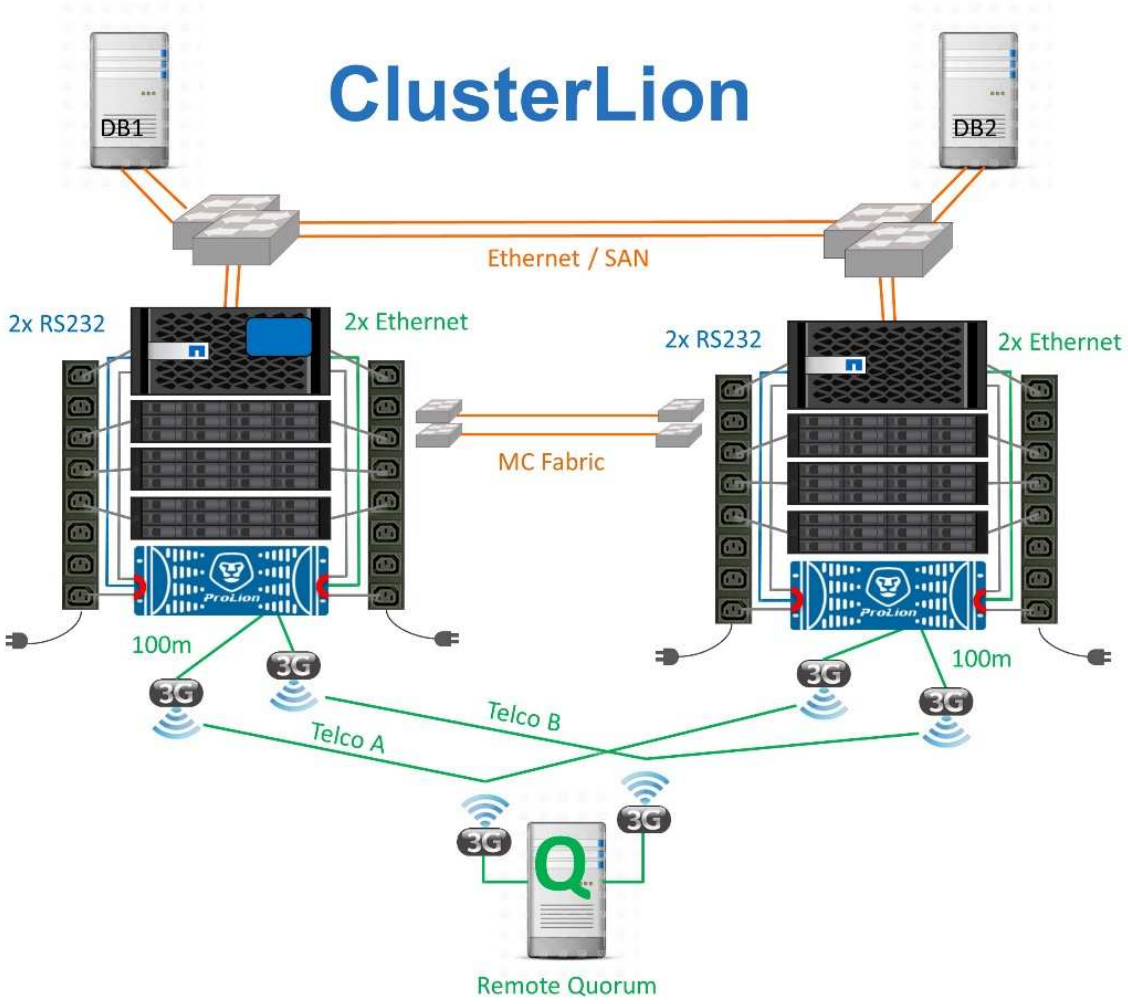
Le médiateur ONTAP est utilisé avec MetroCluster IP et certaines autres solutions ONTAP. Il fonctionne comme un service disjoncteur d'attache classique, tout comme le logiciel disjoncteur d'attache MetroCluster mentionné ci-dessus, mais comprend également une fonctionnalité essentielle, qui effectue un basculement automatique sans surveillance.

Un MetroCluster FAS dispose d'un accès direct aux dispositifs de stockage sur le site opposé. Cela permet à un contrôleur MetroCluster de surveiller l'intégrité des autres contrôleurs en lisant les données de pulsation à partir des disques. Cela permet à un contrôleur de reconnaître la défaillance d'un autre contrôleur et d'effectuer un basculement.

En revanche, l'architecture IP MetroCluster achemine toutes les E/S exclusivement via la connexion contrôleur-contrôleur ; il n'y a pas d'accès direct aux dispositifs de stockage sur le site distant. Cela limite la capacité d'un contrôleur à détecter les défaillances et à effectuer un basculement. Le Mediator ONTAP est donc requis comme dispositif Tiebreaker pour détecter la perte du site et effectuer automatiquement un basculement.

Troisième site virtuel avec ClusterLion

ClusterLion est un dispositif de surveillance MetroCluster avancé qui fonctionne comme un troisième site virtuel. Cette approche permet de déployer MetroCluster en toute sécurité dans une configuration à deux sites avec une fonctionnalité de basculement entièrement automatisée. De plus, ClusterLion peut effectuer un moniteur de niveau réseau supplémentaire et exécuter des opérations de post-basculement. La documentation complète est disponible auprès de ProLion.



- Les appliances ClusterLion contrôlent l'état des contrôleurs à l'aide de câbles série et Ethernet directement connectés.
- Les deux appareils sont connectés l'un à l'autre à l'aide de connexions 3G sans fil redondantes.
- L'alimentation vers le contrôleur ONTAP est acheminée via des relais internes. En cas de panne de site, ClusterLion, qui contient un système UPS interne, coupe les connexions d'alimentation avant d'appeler un basculement. Ce processus permet de s'assurer qu'aucune condition de split-brain ne se produit.
- ClusterLion effectue un basculement dans le délai d'attente SyncMirror de 30 secondes ou pas du tout.

- ClusterLion n’effectue pas de basculement à moins que les États des plexes NVRAM et SyncMirror ne soient synchronisés.
- Étant donné que ClusterLion effectue un basculement uniquement si MetroCluster est entièrement synchronisé, NVFAIL n’est pas nécessaire. Cette configuration permet aux environnements couvrant l’ensemble des sites, tels qu’un RAC Oracle étendu, de rester en ligne, même pendant un basculement non planifié.
- Il inclut les protocoles Fabric-Attached MetroCluster et MetroCluster IP

SyncMirror

Le socle de la protection des données Oracle avec un système MetroCluster est SyncMirror, une technologie de mise en miroir synchrone scale-out aux performances maximales.

Protection des données avec SyncMirror

Au niveau le plus simple, la réplication synchrone implique que toute modification doit être apportée des deux côtés du stockage en miroir avant d’être reconnue. Par exemple, si une base de données écrit un journal ou si un invité VMware est en cours de correction, une écriture ne doit jamais être perdue. Au niveau du protocole, le système de stockage ne doit pas accuser réception de l’écriture tant qu’il n’a pas été validé sur un support non volatile des deux sites. Ce n’est qu’à cette condition qu’il est possible de continuer sans risque de perte de données.

L’utilisation d’une technologie de réplication synchrone est la première étape de la conception et de la gestion d’une solution de réplication synchrone. Il est important de comprendre ce qui pourrait se passer lors de divers scénarios de défaillance planifiés ou non. Les solutions de réplication synchrone offrent toutes des fonctionnalités différentes. Si vous avez besoin d’une solution avec un objectif de point de récupération de zéro, c’est-à-dire sans perte de données, tous les scénarios de défaillance doivent être pris en compte. En particulier, quel est le résultat escompté lorsque la réplication est impossible en raison d’une perte de connectivité entre les sites ?

Disponibilité des données SyncMirror

La réplication MetroCluster repose sur la technologie NetApp SyncMirror, conçue pour basculer efficacement en mode synchrone et en sortir. Cette fonctionnalité répond aux exigences des clients qui demandent une réplication synchrone, mais qui ont également besoin d’une haute disponibilité pour leurs services de données. Par exemple, si la connectivité à un site distant est coupée, il est généralement préférable que le système de stockage continue de fonctionner dans un état non répliqué.

De nombreuses solutions de réplication synchrone ne peuvent fonctionner qu’en mode synchrone. Ce type de réplication « tout ou rien » est parfois appelé mode domino. Ces systèmes de stockage cessent d’accéder aux données au lieu d’interrompre la synchronisation des copies locales et distantes des données. Si la réplication est forcée, la resynchronisation peut prendre beaucoup de temps et laisser un client exposé à des pertes de données complètes pendant la période de rétablissement de la mise en miroir.

Non seulement SyncMirror peut basculer en mode synchrone sans interruption si le site distant est inaccessible, mais il peut également rapidement resynchroniser vers un état RPO = 0 une fois la connectivité restaurée. La copie obsolète des données sur le site distant peut également être conservée dans un état utilisable lors de la resynchronisation, garantissant la présence à tout moment de copies locales et distantes des données.

Si le mode domino est requis, NetApp propose SnapMirror synchrone (SM-S). Des options au niveau de l’application existent également, comme Oracle DataGuard ou SQL Server Always On Availability Groups. La

mise en miroir des disques au niveau du système d'exploitation peut être optionnelle. Pour plus d'informations et d'options, consultez votre équipe de compte NetApp ou partenaire.

MetroCluster et NVFAIL

NVFAIL est une fonctionnalité d'intégrité générale des données de ONTAP conçue pour optimiser la protection de l'intégrité des données avec les bases de données.

> Cette section décrit en détail les fonctionnalités de base de ONTAP NVFAIL et aborde également les sujets spécifiques à MetroCluster.

Avec MetroCluster, une écriture n'est pas confirmée tant qu'elle n'a pas été connectée à la NVRAM et à la NVRAM locales sur au moins un autre contrôleur. Cette approche évite toute panne matérielle ou de courant qui entraîne une perte des E/S à la volée En cas de panne de la mémoire NVRAM locale ou de la connectivité aux autres nœuds, les données ne seront plus mises en miroir.

Si la mémoire NVRAM locale signale une erreur, le nœud s'arrête. Cet arrêt entraîne le basculement vers un contrôleur partenaire lorsque des paires haute disponibilité sont utilisées. Avec MetroCluster, le comportement dépend de la configuration globale choisie, mais il peut entraîner un basculement automatique vers la note distante. Dans tous les cas, aucune donnée n'est perdue parce que le contrôleur qui connaît la défaillance n'a pas acquitté l'opération d'écriture.

Une défaillance de connectivité site à site qui bloque la réplication NVRAM sur des nœuds distants est une situation plus compliquée. Les écritures ne sont plus répliquées sur les nœuds distants, ce qui crée un risque de perte de données en cas d'erreur catastrophique sur un contrôleur. Plus important encore, une tentative de basculement vers un autre nœud dans ces conditions entraîne une perte de données.

Le facteur de contrôle est de savoir si la NVRAM est synchronisée. Si la mémoire NVRAM est synchronisée, le basculement nœud à nœud peut se poursuivre sans risque de perte de données. Dans une configuration MetroCluster, si la mémoire NVRAM et les plexes d'agrégats sous-jacents sont synchronisés, vous pouvez effectuer le basculement sans risque de perte de données.

ONTAP n'autorise pas le basculement ou le basculement lorsque les données ne sont pas synchronisées, sauf si le basculement ou le basculement est forcé. Le fait de forcer une modification des conditions de cette manière reconnaît que les données peuvent être laissées pour compte dans le contrôleur d'origine et que la perte de données est acceptable.

Les bases de données sont particulièrement vulnérables à la corruption si un basculement ou un basculement est forcé, car les bases de données conservent des caches internes de données plus volumineux sur disque. En cas de basculement forcé ou de basculement forcé, les modifications précédemment reconnues sont effectivement supprimées. Le contenu de la baie de stockage recule dans le temps et l'état du cache de la base de données ne reflète plus l'état des données sur le disque.

Afin de protéger les applications de cette situation, ONTAP permet de configurer les volumes pour une protection spéciale contre les défaillances de mémoire NVRAM. Lorsqu'il est déclenché, ce mécanisme de protection entraîne l'entrée d'un volume dans un état appelé NVFAIL. Cet état entraîne des erreurs d'E/S qui entraînent l'arrêt d'une application et n'utilisent donc pas de données obsolètes. Les données ne doivent pas être perdues car des écritures reconnues sont toujours présentes sur le système de stockage et, avec les bases de données, toutes les données de transaction validées doivent être présentes dans les journaux.

Les étapes suivantes habituelles sont qu'un administrateur arrête complètement les hôtes avant de remettre manuellement en ligne les LUN et les volumes. Bien que ces étapes puissent impliquer un certain travail, cette approche est le moyen le plus sûr d'assurer l'intégrité des données. Toutes les données n'ont pas besoin de cette protection. C'est pourquoi NVFAIL peut être configuré volume par volume.

NVFAIL forcé manuellement

Pour forcer un basculement avec un cluster d'applications (y compris VMware, Oracle RAC et autres) distribué sur plusieurs sites, il faut spécifier la méthode la plus sûre `-force-nvfail-all` en ligne de commande. Cette option est disponible en tant que mesure d'urgence pour s'assurer que toutes les données mises en cache sont vidées. Si un hôte utilise des ressources de stockage initialement situées sur le site sinistré, il reçoit des erreurs d'E/S ou un descripteur de fichier obsolète (`ESTALE`) erreur. Les bases de données Oracle planent et les systèmes de fichiers passent entièrement hors ligne ou en mode lecture seule.

Une fois le basculement terminé, le `in-nvfailed-state` L'indicateur doit être effacé et les LUN doivent être mis en ligne. Une fois cette activité terminée, la base de données peut être redémarrée. Ces tâches peuvent être automatisées afin de réduire le RTO.

dr-force-nvfail

En tant que mesure de sécurité générale, réglez le `dr-force-nvfail` drapeau sur tous les volumes accessibles depuis un site distant pendant les opérations normales, ce qui signifie qu'il s'agit d'activités utilisées avant le basculement. Le résultat de ce paramètre est que les volumes distants sélectionnés deviennent indisponibles lorsqu'ils entrent `in-nvfailed-state` lors d'un basculement. Une fois le basculement terminé, le `in-nvfailed-state` L'indicateur doit être effacé et les LUN doivent être mis en ligne. Une fois ces activités terminées, les applications peuvent être redémarrées. Ces tâches peuvent être automatisées afin de réduire le RTO.

Le résultat est similaire à l'utilisation du `-force-nvfail-all` indicateur pour commutateurs manuels. Toutefois, le nombre de volumes affectés peut être limité aux volumes qui doivent être protégés contre les applications ou les systèmes d'exploitation dotés de caches obsolètes.

> Il existe deux exigences critiques pour un environnement qui n'utilise pas `dr-force-nvfail` sur les volumes d'application :

- Un basculement forcé ne doit pas se produire plus de 30 secondes après la perte du site principal.
- Le basculement ne doit pas avoir lieu pendant les tâches de maintenance ou tout autre mode dans lequel les plexes SyncMirror ou la réplication NVRAM sont désynchronisés. Le premier critère peut être atteint à l'aide d'un logiciel disjoncteur d'attache configuré pour effectuer un basculement dans les 30 secondes qui suivent la défaillance d'un site. Cela ne signifie pas que le basculement doit être effectué dans les 30 secondes qui suivent la détection d'une défaillance de site. Cela signifie qu'il n'est plus sûr de forcer un basculement si 30 secondes se sont écoulées depuis qu'un site a été confirmé opérationnel.

Le deuxième critère peut être partiellement respecté en désactivant toutes les fonctionnalités de basculement automatisé lorsque la configuration MetroCluster est désynchronisée. Il est préférable d'opter pour une solution disjoncteur d'attache capable de surveiller l'état de santé de la réplication NVRAM et des plexes SyncMirror. Si le cluster n'est pas entièrement synchronisé, le disjoncteur d'attache ne doit pas déclencher de basculement.

Le logiciel MCTB de NetApp ne peut pas contrôler l'état de la synchronisation. Il doit donc être désactivé lorsque MetroCluster n'est pas synchronisé pour quelque raison que ce soit. ClusterLion inclut des fonctionnalités de surveillance NVRAM et plex et peut être configuré pour ne pas déclencher le basculement à moins que le système MetroCluster ne soit entièrement synchronisé.

Instance unique Oracle

Comme indiqué précédemment, la présence d'un système MetroCluster n'ajoute pas nécessairement aux meilleures pratiques d'exploitation d'une base de données ou ne les

modifie pas nécessairement. La majorité des bases de données qui s'exécutent actuellement sur les systèmes MetroCluster client sont à instance unique et suivent les recommandations de la documentation Oracle sur ONTAP.

Basculement avec un système d'exploitation préconfiguré

SyncMirror livre une copie synchrone des données au niveau du site de reprise d'activité. La mise à disposition des données requiert un système d'exploitation et les applications associées. L'automatisation de base peut considérablement améliorer le délai de basculement de l'environnement global. Les produits Clusterware tels que Veritas Cluster Server (VCS) sont souvent utilisés pour créer un cluster sur les sites et, dans la plupart des cas, le processus de basculement peut être piloté par des scripts simples.

En cas de perte des nœuds principaux, le cluster (ou les scripts) est configuré de manière à mettre les bases de données en ligne sur le site secondaire. Une option consiste à créer des serveurs de secours préconfigurés pour les ressources NFS ou SAN qui constituent la base de données. En cas de défaillance du site principal, le logiciel de mise en cluster ou l'alternative scriptée effectue une séquence d'actions similaires à celles décrites ci-dessous :

1. Forçage du basculement MetroCluster
2. Découverte de LUN FC (SAN uniquement)
3. Montage de systèmes de fichiers et/ou montage de groupes de disques ASM
4. Démarrage de la base de données

Cette approche doit avant tout se passer d'un système d'exploitation en cours d'exécution sur le site distant. Elles doivent être préconfigurées avec des binaires Oracle, ce qui signifie également que des tâches telles que l'application de correctifs Oracle doivent être effectuées sur les sites principal et de secours. Les binaires Oracle peuvent également être mis en miroir vers le site distant et montés en cas d'incident.

La procédure d'activation réelle est simple. Les commandes telles que la découverte de LUN ne nécessitent que quelques commandes par port FC. Le montage du système de fichiers n'est rien de plus qu'un `mount` Et les bases de données et ASM peuvent être démarrés et arrêtés sur l'interface de ligne de commande à l'aide d'une seule commande. Si les volumes et les systèmes de fichiers ne sont pas utilisés sur le site de reprise d'activité avant le basculement, il n'est pas nécessaire de les définir `dr-force- nvfail` sur les volumes.

Basculement avec un système d'exploitation virtualisé

Le basculement des environnements de base de données peut être étendu pour inclure le système d'exploitation lui-même. En théorie, ce basculement peut être effectué avec des LUN de démarrage, mais le plus souvent avec un système d'exploitation virtualisé. La procédure est similaire aux étapes suivantes :

1. Forçage du basculement MetroCluster
2. Montage des datastores hébergeant les machines virtuelles du serveur de base de données
3. Démarrage des machines virtuelles
4. Démarrage manuel des bases de données ou configuration des machines virtuelles pour démarrer automatiquement les bases de données par exemple, un cluster ESX peut couvrir des sites. En cas d'incident, les machines virtuelles peuvent être mises en ligne sur le site de reprise après incident après le basculement. Tant que les datastores hébergeant les serveurs de base de données virtualisés ne sont pas utilisés au moment de l'incident, il n'est pas nécessaire de les définir `dr-force- nvfail` sur les volumes associés.

RAC étendu Oracle

De nombreux clients optimisent leur RTO en étendant un cluster Oracle RAC sur plusieurs sites, offrant une configuration entièrement active/active. La conception globale devient plus complexe car elle doit inclure la gestion du quorum d'Oracle RAC. En outre, l'accès aux données se fait depuis les deux sites, ce qui signifie qu'un basculement forcé peut entraîner l'utilisation d'une copie obsolète des données.

Bien qu'une copie des données soit présente sur les deux sites, seul le contrôleur qui possède actuellement un agrégat peut assurer le service des données. Par conséquent, avec les clusters RAC étendus, les nœuds distants doivent effectuer des E/S sur une connexion site à site. Il en résulte une latence d'E/S supplémentaire, mais cette latence n'est généralement pas problématique. Le réseau d'interconnexion RAC doit également être étendu entre les sites, ce qui signifie qu'un réseau haut débit à faible latence est requis de toute façon. Si la latence supplémentaire pose problème, le cluster peut être exploité de manière actif-passif. Les opérations exigeantes en E/S devront ensuite être dirigées vers les nœuds RAC locaux vers le contrôleur propriétaire des agrégats. Les nœuds distants effectuent alors des opérations d'E/S plus légères ou sont utilisés uniquement comme serveurs de secours.

Si un RAC étendu actif-actif est requis, la synchronisation active SnapMirror doit être considérée à la place de MetroCluster. La réplication SM-AS permet de privilégier une réplique spécifique des données. Par conséquent, un cluster RAC étendu peut être intégré dans lequel toutes les lectures se produisent localement. Les E/S de lecture ne traversent jamais les sites, ce qui assure la latence la plus faible possible. Toute activité d'écriture doit toujours transiter la connexion intersite, mais ce trafic est inévitable avec toute solution de mise en miroir synchrone.

> Si des LUN de démarrage, y compris des disques de démarrage virtualisés, sont utilisés avec Oracle RAC, il `misscount` peut être nécessaire de modifier le paramètre. Pour plus d'informations sur les paramètres de délai d'expiration du RAC, reportez-vous à la section "Oracle RAC avec ONTAP".

Configuration à deux sites

Une configuration RAC étendue sur deux sites peut fournir des services de base de données actif-actif qui peuvent survivre à de nombreux scénarios d'incident, mais pas à tous, sans interruption.

Fichiers de vote RAC

La gestion du quorum doit être prise en compte lors du déploiement du RAC étendu sur MetroCluster. Oracle RAC dispose de deux mécanismes pour gérer le quorum : le battement de cœur du disque et le battement de cœur du réseau. La pulsation du disque surveille l'accès au stockage à l'aide des fichiers de vote. Dans le cas d'une configuration RAC à site unique, une ressource de vote unique suffit tant que le système de stockage sous-jacent offre des fonctionnalités haute disponibilité.

Dans les versions précédentes d'Oracle, les fichiers de vote étaient placés sur des périphériques de stockage physiques, mais dans les versions actuelles d'Oracle, les fichiers de vote sont stockés dans des groupes de disques ASM.

> Oracle RAC est pris en charge par NFS. Pendant le processus d'installation de la grille, un ensemble de processus ASM est créé pour présenter l'emplacement NFS utilisé pour les fichiers de grille en tant que groupe de disques ASM. Le processus est presque transparent pour l'utilisateur final et ne nécessite aucune gestion ASM continue une fois l'installation terminée.

Dans une configuration à deux sites, il est tout d'abord nécessaire de s'assurer que chaque site peut toujours accéder à plus de la moitié des fichiers de vote, ce qui garantit un processus de reprise après incident sans interruption. Cette tâche était simple avant que les fichiers de vote ne soient stockés dans des groupes de disques ASM, mais aujourd'hui, les administrateurs doivent comprendre les principes de base de la redondance ASM.

Les groupes de disques ASM disposent de trois options de redondance `external`, `normal`, et `high`. En d'autres termes, sans miroir, avec miroir et miroir à 3 voies. Une option plus récente appelée `Flex` est également disponible, mais rarement utilisé. Le niveau de redondance et le placement des périphériques redondants contrôlent ce qui se passe dans les scénarios de panne. Par exemple :

- Placer les fichiers de vote sur un `diskgroup` avec `external` la redondance des ressources garantit la suppression d'un site en cas de perte de la connectivité intersite.
- Placer les fichiers de vote sur un `diskgroup` avec `normal` La redondance avec un seul disque ASM par site garantit la suppression des nœuds sur les deux sites en cas de perte de la connectivité intersite, car aucun des sites ne possède un quorum majoritaire.
- Placer les fichiers de vote sur un `diskgroup` avec `high` la redondance avec deux disques sur un site et un seul disque sur l'autre site permet des opérations actif-actif lorsque les deux sites sont opérationnels et mutuellement accessibles. Toutefois, si le site à disque unique est isolé du réseau, ce site est supprimé.

Pulsation du réseau RAC

Le signal de présence du réseau RAC Oracle surveille l'accessibilité des nœuds sur l'interconnexion de cluster. Pour rester dans le cluster, un nœud doit pouvoir contacter plus de la moitié des autres nœuds. Dans une architecture à deux sites, cette exigence crée les choix suivants pour le nombre de nœuds RAC :

- Le placement d'un nombre égal de nœuds par site entraîne la suppression sur un site en cas de perte de la connectivité réseau.
- Le placement de N nœuds sur un site et de N+1 nœuds sur le site opposé garantit que la perte de la connectivité intersite entraîne le site avec le plus grand nombre de nœuds restants dans le quorum du réseau et le site avec moins de nœuds supprimés.

Avant Oracle 12cR2, il était impossible de contrôler quel côté devait être expulsé en cas de perte du site. Lorsque chaque site a un nombre égal de nœuds, l'exclusion est contrôlée par le nœud maître, qui est en général le premier nœud RAC à démarrer.

Oracle 12cR2 introduit la fonctionnalité de pondération des nœuds. L'administrateur peut ainsi mieux contrôler la manière dont Oracle résout les problèmes de partage du cerveau. À titre d'exemple simple, la commande suivante définit les préférences pour un nœud particulier dans un RAC :

```
[root@host-a ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.
```

Après le redémarrage d'Oracle High-Availability Services, la configuration se présente comme suit :

```
[root@host-a lib]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
```

Nœud `host-a` est maintenant désigné comme serveur critique. Si les deux nœuds RAC sont isolés, `host-a` survit, et `host-b` est supprimé.

> Pour plus d'informations, consultez le livre blanc Oracle « Oracle Clusterware 12c Release 2 Technical Overview. "

Pour les versions d'Oracle RAC antérieures à 12cR2, le nœud maître peut être identifié en vérifiant les journaux CRS comme suit :

```
[root@host-a ~]# /grid/bin/crsctl status server -f | egrep
'^NAME|CSS_CRITICAL='
NAME=host-a
CSS_CRITICAL=yes
NAME=host-b
CSS_CRITICAL=no
 [root@host-a ~]# grep -i 'master node' /grid/diag/crs/host-
a/crs/trace/crsd.trc
2017-05-04 04:46:12.261525 :   CRSSE:2130671360: {1:16377:2} Master Change
Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:01:24.979716 :   CRSSE:2031576832: {1:13237:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
2017-05-04 05:11:22.995707 :   CRSSE:2031576832: {1:13237:221} Master
Change Event; New Master Node ID:1 This Node's ID:1
2017-05-04 05:28:25.797860 :   CRSSE:3336529664: {1:8557:2} Master Change
Event; New Master Node ID:2 This Node's ID:1
```

Ce journal indique que le nœud maître est `2` et le nœud `host-a` A un ID de `1`. Ce fait signifie que `host-a` n'est pas le nœud maître. L'identité du nœud maître peut être confirmée avec la commande `olsnodes -n`.

```
[root@host-a ~]# /grid/bin/olsnodes -n
host-a  1
host-b  2
```

Le nœud ayant l'ID de `2` est `host-b`, qui est le nœud maître. Dans une configuration avec un nombre égal de nœuds sur chaque site, le site avec `host-b` est le site qui survit si les deux ensembles perdent la connectivité réseau pour quelque raison que ce soit.

Il est possible que l'entrée de journal qui identifie le nœud maître puisse sortir du système. Dans ce cas, les horodatages des sauvegardes du registre des clusters Oracle (OCR) peuvent être utilisés.

```
[root@host-a ~]#  /grid/bin/ocrconfig -showbackup
host-b     2017/05/05 05:39:53     /grid/cdata/host-cluster/backup00.ocr
0
host-b     2017/05/05 01:39:53     /grid/cdata/host-cluster/backup01.ocr
0
host-b     2017/05/04 21:39:52     /grid/cdata/host-cluster/backup02.ocr
0
host-a     2017/05/04 02:05:36     /grid/cdata/host-cluster/day.ocr     0
host-a     2017/04/22 02:05:17     /grid/cdata/host-cluster/week.ocr     0
```

Cet exemple montre que le nœud maître est `host-b`. Il indique également un changement dans le nœud maître de `host-a` à `host-b` Quelque part entre 2:05 et 21:39 le 4 mai. Cette méthode d'identification du nœud maître n'est sûre que si les journaux CRS ont également été vérifiés car il est possible que le nœud maître ait changé depuis la sauvegarde OCR précédente. Si ce changement s'est produit, il doit être visible dans les journaux OCR.

La plupart des clients choisissent un seul groupe de disques de vote qui dessert l'ensemble de l'environnement et un nombre égal de nœuds RAC sur chaque site. Le groupe de disques doit être placé sur le site qui contient la base de données. En conséquence, une perte de connectivité entraîne la suppression du site distant. Le site distant n'aurait plus le quorum, ni l'accès aux fichiers de base de données, mais le site local continue à fonctionner normalement. Une fois la connectivité rétablie, l'instance distante peut être de nouveau mise en ligne.

En cas d'incident, un basculement est nécessaire pour mettre en ligne les fichiers de base de données et le groupe de disques de vote sur le site survivant. Si l'incident permet à AUSO de déclencher le basculement, NVFAIL n'est pas déclenché, car le cluster est connu pour être synchronisé et les ressources de stockage sont normalement mises en ligne. L'AUSO est une opération très rapide et doit se terminer avant le `disktimeout` la période expire.

Comme il n'y a que deux sites, il n'est pas possible d'utiliser n'importe quel type de logiciel automatisé externe de rupture de tieBreaking, ce qui signifie que le basculement forcé doit être une opération manuelle.

Configurations à trois sites

Un cluster RAC étendu est beaucoup plus facile à concevoir avec trois sites. Les deux sites hébergeant chaque moitié du système MetroCluster prennent également en charge les workloads de la base de données, tandis que le troisième sert de disjoncteur pour la base de données et le système MetroCluster. La configuration Oracle Tiebreaker peut être aussi simple que le placement d'un membre du groupe de disques ASM utilisé pour le vote sur un troisième site, et peut également inclure une instance opérationnelle sur le troisième site pour s'assurer qu'il y a un nombre impair de nœuds dans le cluster RAC.

> Consultez la documentation Oracle sur « quorum failure group » pour obtenir des informations importantes sur l'utilisation de NFS dans une configuration RAC étendue. En résumé, il peut être nécessaire de modifier les options de montage NFS pour inclure l'option logicielle permettant de s'assurer que la perte de connectivité au troisième site hébergeant les ressources quorum n'affecte pas les serveurs Oracle ou les processus RAC Oracle principaux.

Synchronisation active SnapMirror

Présentation

La synchronisation active SnapMirror vous permet de créer des environnements de base de données Oracle à ultra haute disponibilité où des LUN sont disponibles à partir de deux clusters de stockage différents.

Avec la synchronisation active SnapMirror, il n'y a pas de copie « principale » ni de copie « secondaire » des données. Chaque cluster peut fournir des E/S de lecture à partir de sa copie locale des données, et chaque cluster réplique une écriture vers son partenaire. Le résultat est un comportement d'E/S symétrique.

Entre autres options, vous pouvez exécuter Oracle RAC en tant que cluster étendu avec des instances opérationnelles sur les deux sites. Vous pouvez également créer des clusters de bases de données actif-passif RPO=0, dans lesquels les bases de données à instance unique peuvent être déplacées entre les sites en cas de panne sur le site. Ce processus peut également être automatisé via des produits tels que Pacemaker ou VMware HA. Toutes ces options reposent sur la réplication synchrone gérée par SnapMirror Active Sync.

Réplication synchrone

En fonctionnement normal, la synchronisation active SnapMirror fournit en permanence une réplique synchrone avec un objectif de point de récupération de 0, à une exception près. Si les données ne peuvent pas être répliquées, ONTAP exige de répliquer les données et de reprendre le traitement des E/S sur un site pendant que les LUN de l'autre site sont mises hors ligne.

Matériel de stockage

Contrairement à d'autres solutions de reprise après incident du stockage, la synchronisation active SnapMirror offre une flexibilité asymétrique de la plateforme. Le matériel de chaque site n'a pas besoin d'être identique. Cette fonctionnalité vous permet d'ajuster la taille du matériel utilisé pour prendre en charge la synchronisation active SnapMirror. Le système de stockage distant peut être identique au site principal s'il doit prendre en charge une charge de travail de production complète, mais si un incident entraîne une réduction des E/S, un système plus petit sur le site distant peut être plus économique.

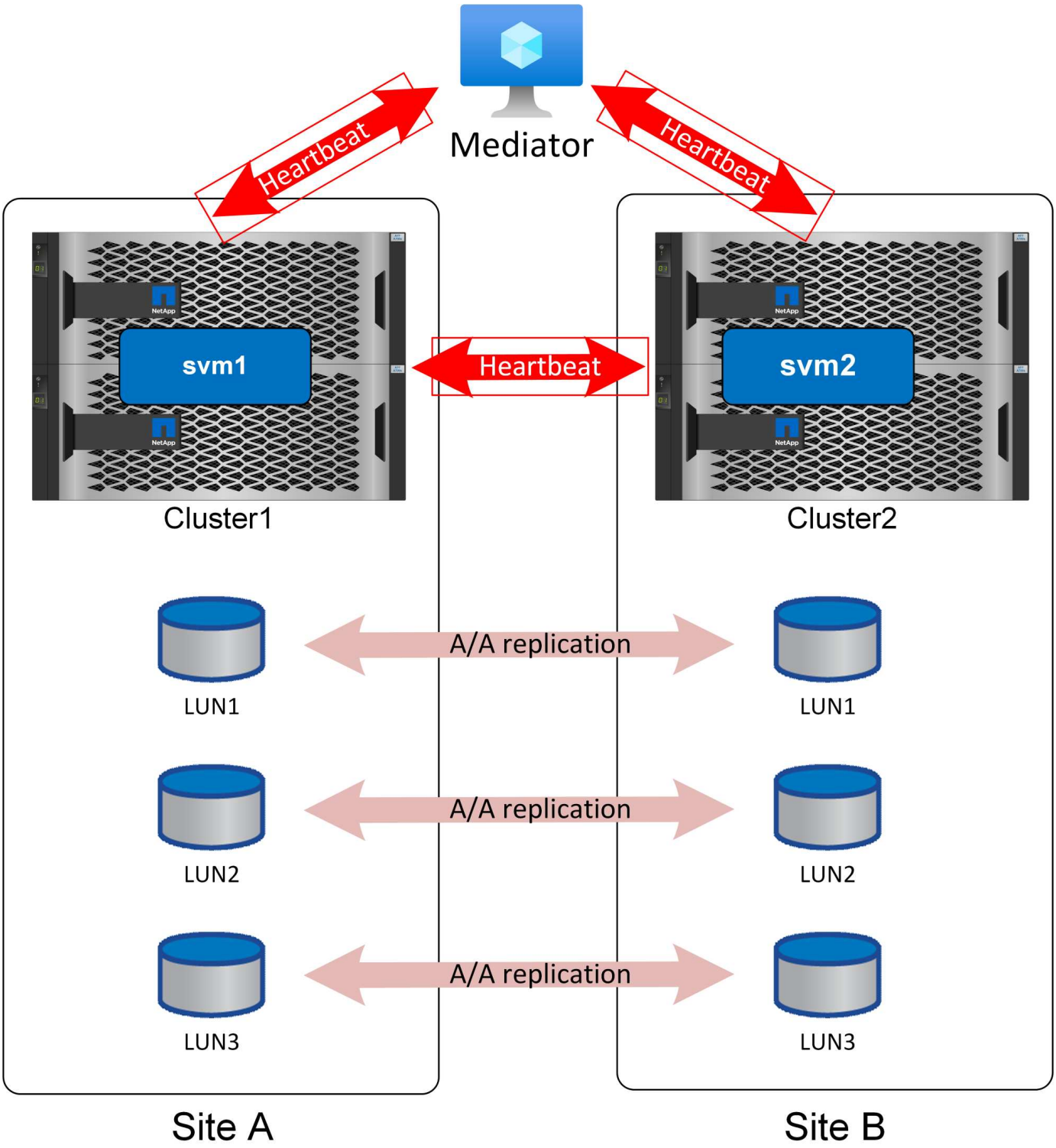
ONTAP Médiateur

Le médiateur ONTAP est une application logicielle téléchargée depuis la prise en charge de NetApp et généralement déployée sur une petite machine virtuelle. Le Mediator ONTAP n'est pas un Tiebreaker lorsqu'il est utilisé avec la synchronisation active SnapMirror. Il s'agit d'un canal de communication alternatif pour les deux clusters qui participent à la réplication SnapMirror active Sync. Les opérations automatisées sont dirigées par ONTAP sur la base des réponses reçues du partenaire via des relations directes et via le médiateur.

Médiateur de ONTAP

Le médiateur est requis pour automatiser le basculement en toute sécurité. Dans l'idéal, elle serait placée sur un site tiers indépendant, mais elle peut toujours fonctionner pour la plupart des besoins si elle est en colocation avec l'un des clusters participant à la réplication.

Le médiateur n'est pas vraiment un casse-barre, bien que c'est effectivement la fonction qu'il fournit. Il ne prend aucune action ; il fournit plutôt un canal de communication alternatif pour la communication cluster à cluster.



Le principal défi lié au basculement automatisé est le problème des réseaux partagés, qui se pose en cas de perte de connectivité entre les deux sites. Que doit-on faire ? Vous ne voulez pas que deux sites différents se désignent comme les copies restantes des données, mais comment un seul site peut-il faire la différence entre la perte réelle du site opposé et l'incapacité à communiquer avec le site opposé ?

C'est là que le médiateur entre dans la photo. S'il est placé sur un troisième site, et chaque site a une connexion réseau distincte à ce site, alors vous avez un chemin supplémentaire pour chaque site pour valider l'état de santé de l'autre. Examinez à nouveau l'image ci-dessus et examinez les scénarios suivants.

- Que se passe-t-il si le médiateur échoue ou est inaccessible à partir d'un ou des deux sites ?
  - Les deux clusters peuvent toujours communiquer entre eux sur le même lien que celui utilisé pour les services de réplication.
  - Les données restent protégées avec un objectif de point de récupération de 0
- Que se passe-t-il si le site A tombe en panne ?
  - Le site B verra les deux canaux de communication tomber en panne.
  - Le site B prendra le contrôle des services de données, mais sans mise en miroir RPO=0
- Que se passe-t-il si le site B tombe en panne ?
  - Le site A verra les deux canaux de communication tomber en panne.
  - Le site A prend le relais des services de données, mais sans mise en miroir avec un objectif de point de récupération de 0

Il y a un autre scénario à prendre en compte : la perte du lien de réplication des données. En cas de perte de la liaison de réplication entre les sites, la mise en miroir avec un objectif de point de récupération de 0 sera évidemment impossible. Que devrait-on alors se passer ?

Ceci est contrôlé par le statut du site préféré. Dans une relation SM-AS, l'un des sites est secondaire à l'autre. Cela n'a aucun effet sur les opérations normales, et tout accès aux données est symétrique. Toutefois, si la réplication est interrompue, le nœud devra être rompu pour reprendre les opérations. Par conséquent, le site privilégié continuera les opérations sans mise en miroir et le site secondaire arrêtera le traitement des E/S jusqu'à ce que la communication de réplication soit restaurée.

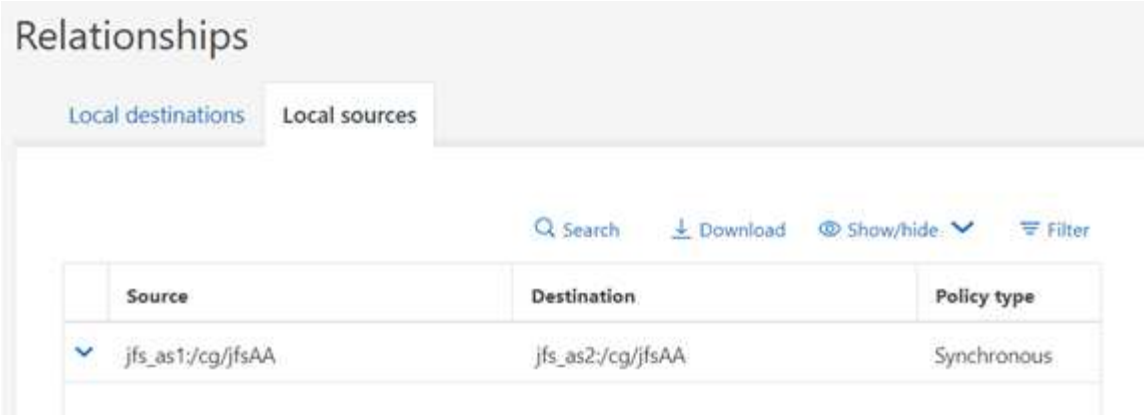
Site préféré de la synchronisation active SnapMirror

Le comportement de la synchronisation active SnapMirror est symétrique, avec une exception importante : la configuration du site préféré.

La synchronisation active SnapMirror considère un site comme la « source » et l'autre comme la « destination ». Cela implique une relation de réplication unidirectionnelle, mais cela ne s'applique pas au comportement d'E/S. La réplication est bidirectionnelle et symétrique. Les temps de réponse d'E/S sont identiques de part et d'autre du miroir.

La `source` désignation est le contrôle du site préféré. En cas de perte du lien de réplication, les chemins de LUN sur la copie source continueront à transmettre des données tandis que les chemins de LUN sur la copie de destination deviendront indisponibles jusqu'à ce que la réplication soit rétablie et que SnapMirror repasse à l'état synchrone. Les chemins reprennent alors le service des données.

La configuration source/destination peut être affichée via SystemManager :



Ou sur l'interface de ligne de commande :

```
Cluster2::> snapmirror show -destination-path jfs_as2:/cg/jfsAA

                           Source Path: jfs_as1:/cg/jfsAA
                      Destination Path: jfs_as2:/cg/jfsAA
                    Relationship Type: XDP
               Relationship Group Type: consistencygroup
                   SnapMirror Schedule: -
                SnapMirror Policy Type: automated-failover-duplex
                     SnapMirror Policy: AutomatedFailOverDuplex
                           Tries Limit: -
                     Throttle (KB/sec): -
                          Mirror State: Snapmirrored
                   Relationship Status: InSync
```

La clé est que la source est le SVM sur le cluster1. Comme mentionné ci-dessus, les termes « source » et « destination » ne décrivent pas le flux des données répliquées. Les deux sites peuvent traiter une écriture et la répliquer sur le site opposé. En effet, les deux grappes sont des sources et des destinations. La désignation d'un cluster comme source contrôle simplement le cluster qui survit en tant que système de stockage en lecture/écriture en cas de perte du lien de réplication.

Topologie réseau

Accès uniforme

Un réseau d'accès uniforme signifie que les hôtes peuvent accéder aux chemins sur les deux sites (ou domaines de défaillance au sein du même site).

L'une des caractéristiques importantes de SM-AS est la capacité de configurer les systèmes de stockage pour savoir où se trouvent les hôtes. Lorsque vous mappez les LUN sur un hôte donné, vous pouvez indiquer si elles sont proximales ou non à un système de stockage donné.

Paramètres de proximité

La proximité fait référence à une configuration par cluster qui indique qu'un WWN d'hôte ou un ID d'initiateur iSCSI appartient à un hôte local. Il s'agit d'une deuxième étape facultative de configuration de l'accès aux

LUN.

La première étape correspond à la configuration habituelle du groupe initiateur. Chaque LUN doit être mappée sur un groupe initiateur qui contient les ID WWN/iSCSI des hôtes devant accéder à cette LUN. Cela contrôle quel hôte a *accès* à un LUN.

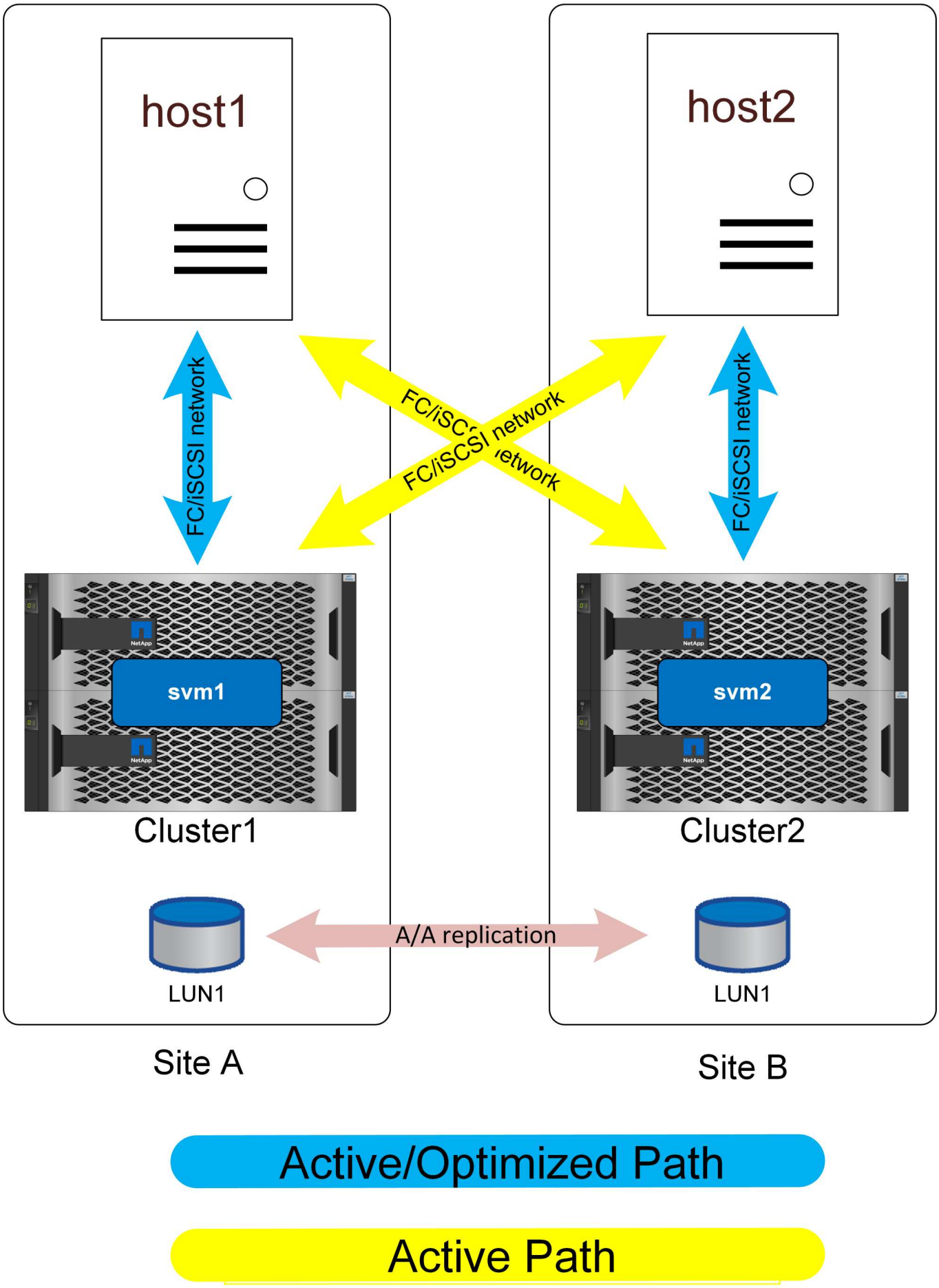
La deuxième étape facultative consiste à configurer la proximité de l'hôte. Cela ne contrôle pas l'accès, il contrôle *Priority*.

Par exemple, un hôte du site A peut être configuré pour accéder à une LUN protégée par la synchronisation active SnapMirror. Le SAN étant étendu entre les sites, les chemins d'accès sont disponibles pour cette LUN via le stockage sur le site A ou le stockage sur le site B.

Sans paramètres de proximité, cet hôte utilisera les deux systèmes de stockage de la même manière, car les deux systèmes de stockage annonceront des chemins actifs/optimisés. Si la latence SAN et/ou la bande passante entre les sites est limitée, il se peut que cela ne soit pas désirable, et vous pouvez vous assurer que, pendant le fonctionnement normal, chaque hôte utilise de préférence des chemins vers le système de stockage local. Cette configuration s'effectue en ajoutant l'ID WWN/iSCSI de l'hôte au cluster local en tant qu'hôte proximal. Cette opération peut être effectuée à partir de l'interface de ligne de commande ou de SystemManager.

AFF

Avec un système AFF, les chemins apparaissent comme indiqué ci-dessous lorsque la proximité de l'hôte a été configurée.


host1
host2
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
svm1
svm2
Cluster1
Cluster2
A/A replication
LUN1
LUN1
Site A
Site B
Active/Optimized Path
Active Path

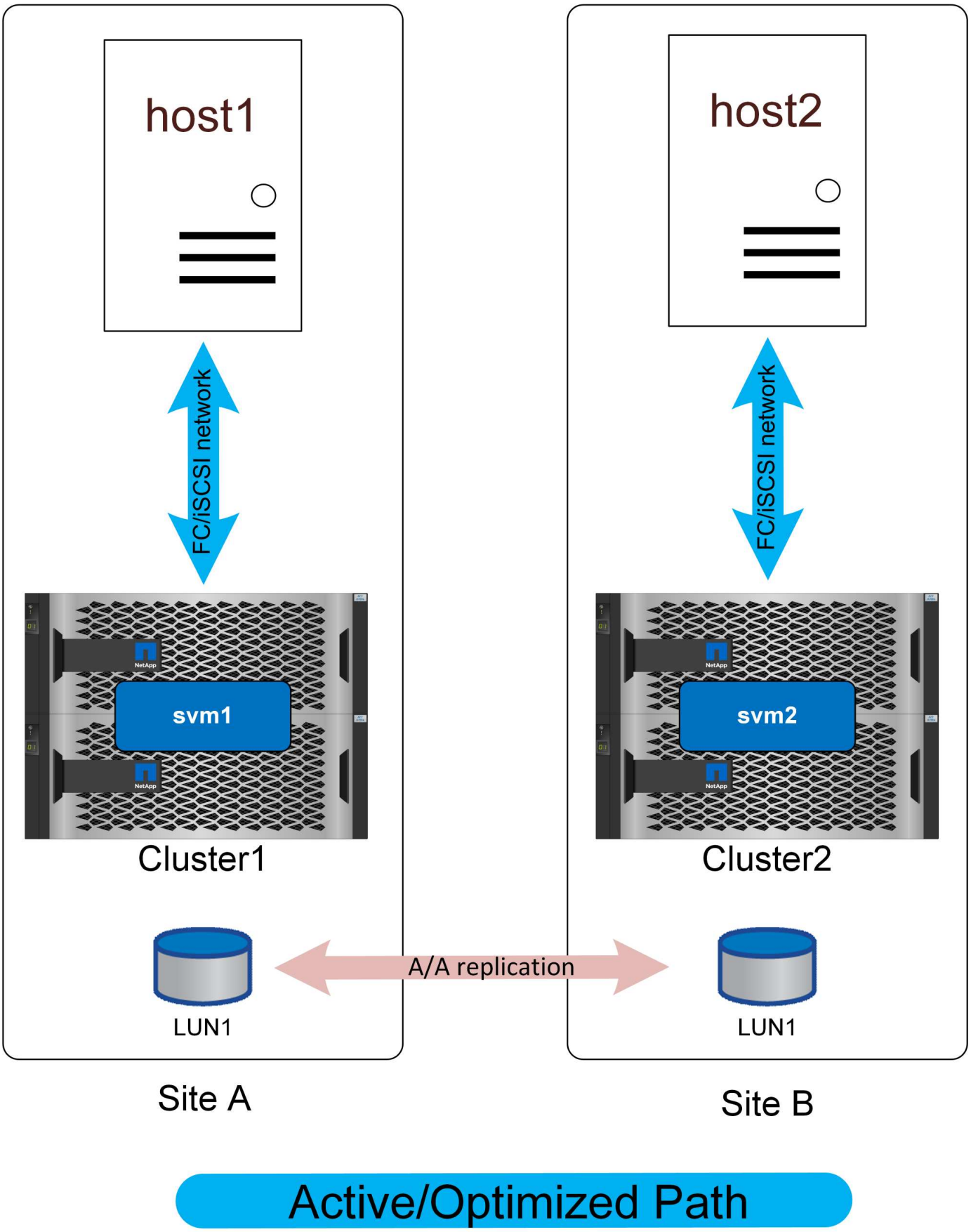
En fonctionnement normal, toutes les E/S sont des E/S locales. Les opérations de lecture et d'écriture sont gérées à partir de la baie de stockage locale. Bien entendu, les E/S en écriture devront également être répliquées par le contrôleur local sur le système distant avant d'être acquittées, mais toutes les E/S en lecture seront gérées localement et ne subiront pas de latence supplémentaire en traversant la liaison SAN entre les sites.

Le seul moment où les chemins non optimisés seront utilisés est la perte de tous les chemins actifs/optimisés. Par exemple, si l'ensemble de la baie sur le site A est hors tension, les hôtes du site A peuvent toujours accéder aux chemins d'accès à la baie sur le site B et donc rester opérationnels, même s'ils connaissent une latence plus élevée.

Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

ASA

Les systèmes NetApp ASA proposent des chemins d'accès multiples actif-actif sur tous les chemins d'accès à un cluster. Cela s'applique également aux configurations SM-AS.



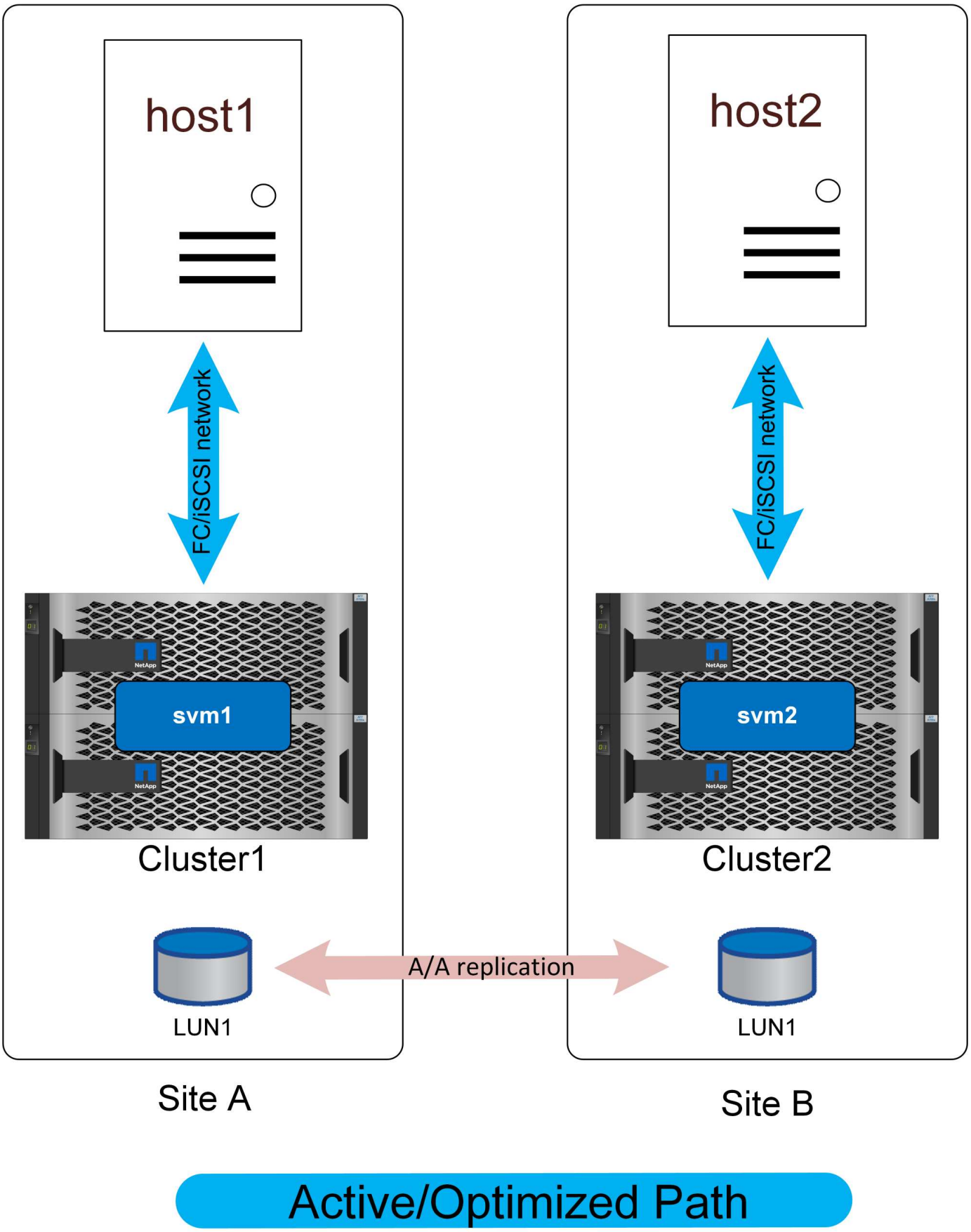
Une configuration ASA avec un accès non uniforme fonctionnera en grande partie comme avec AFF. Avec un accès uniforme, l'E/S traverserait le WAN. Cela peut être souhaitable ou non.

Si les deux sites étaient distants de 100 mètres avec une connectivité à fibre optique, il ne devrait pas y avoir de latence supplémentaire détectable traversant le WAN, mais si les sites étaient éloignés, les performances de lecture seraient affectées sur les deux sites. À l'inverse, avec AFF, ces chemins WAN seraient utilisés uniquement s'il n'existait aucun chemin local disponible et si les performances quotidiennes seraient meilleures, car toutes les E/S seraient des E/S locales. ASA avec un réseau d'accès non uniforme serait une option pour bénéficier des avantages de ASA en termes de coûts et de fonctionnalités sans engendrer de pénalités de latence entre les sites.

ASA avec SM-as dans une configuration à faible latence offre deux avantages intéressants. Tout d'abord, elle double *les performances de n'importe quel hôte, car les E/S peuvent être traitées par deux fois plus de contrôleurs en utilisant deux fois plus de chemins. Ensuite, dans un environnement à site unique, elle offre une disponibilité extrême, car l'intégralité du système de stockage peut être perdue sans interrompre l'accès aux hôtes.

Accès non uniforme

La mise en réseau à accès non uniforme signifie que chaque hôte n'a accès qu'aux ports du système de stockage local. Le SAN n'est pas étendu sur les sites (ou les domaines de défaillance au sein du même site).



Le principal avantage de cette approche est la simplicité du SAN : vous n'avez plus besoin d'étendre un SAN sur le réseau. Certains clients ne disposent pas d'une connectivité à faible latence suffisante entre les sites, ou

n’ont pas l’infrastructure nécessaire pour acheminer le trafic SAN FC sur un réseau intersite.

L’inconvénient de l’accès non uniforme est que certains scénarios de défaillance, notamment la perte du lien de réplication, entraînent la perte de l’accès au stockage par certains hôtes. En cas de perte de la connectivité du stockage local, les applications qui s’exécutent en tant qu’instances uniques, telles qu’une base de données non en cluster et qui ne s’exécute intrinsèquement que sur un hôte unique sur un montage donné, échouent. Les données seraient toujours protégées, mais le serveur de base de données n’aurait plus accès. Il doit être redémarré sur un site distant, de préférence par le biais d’un processus automatisé. Par exemple, VMware HA peut détecter une situation de tous les chemins d’accès sur un serveur et redémarrer une machine virtuelle sur un autre serveur sur lequel les chemins d’accès sont disponibles.

En revanche, une application en cluster telle qu’Oracle RAC peut fournir un service qui est disponible simultanément sur deux sites différents. La perte d’un site ne signifie pas la perte du service d’application dans son ensemble. Les instances restent disponibles et s’exécutent sur le site survivant.

Dans de nombreux cas, la surcharge liée à la latence supplémentaire qu’une application accède au système de stockage via une liaison site à site ne serait pas acceptable. Cela signifie que l’amélioration de la disponibilité des réseaux uniformes est minime, car la perte de stockage sur un site entraînerait la nécessité de fermer les services sur ce site défaillant.

> Il existe des chemins redondants à travers le cluster local qui ne sont pas illustrés sur ces schémas pour plus de simplicité. Les systèmes de stockage ONTAP étant dotés de la haute disponibilité, une panne du contrôleur ne devrait pas entraîner de panne sur le site. Il devrait simplement entraîner une modification dans laquelle les chemins locaux sont utilisés sur le site affecté.

Configurations Oracle

Présentation

L’utilisation de la synchronisation active SnapMirror n’ajoute pas nécessairement aux meilleures pratiques d’exploitation d’une base de données ou ne modifie pas nécessairement ces pratiques.

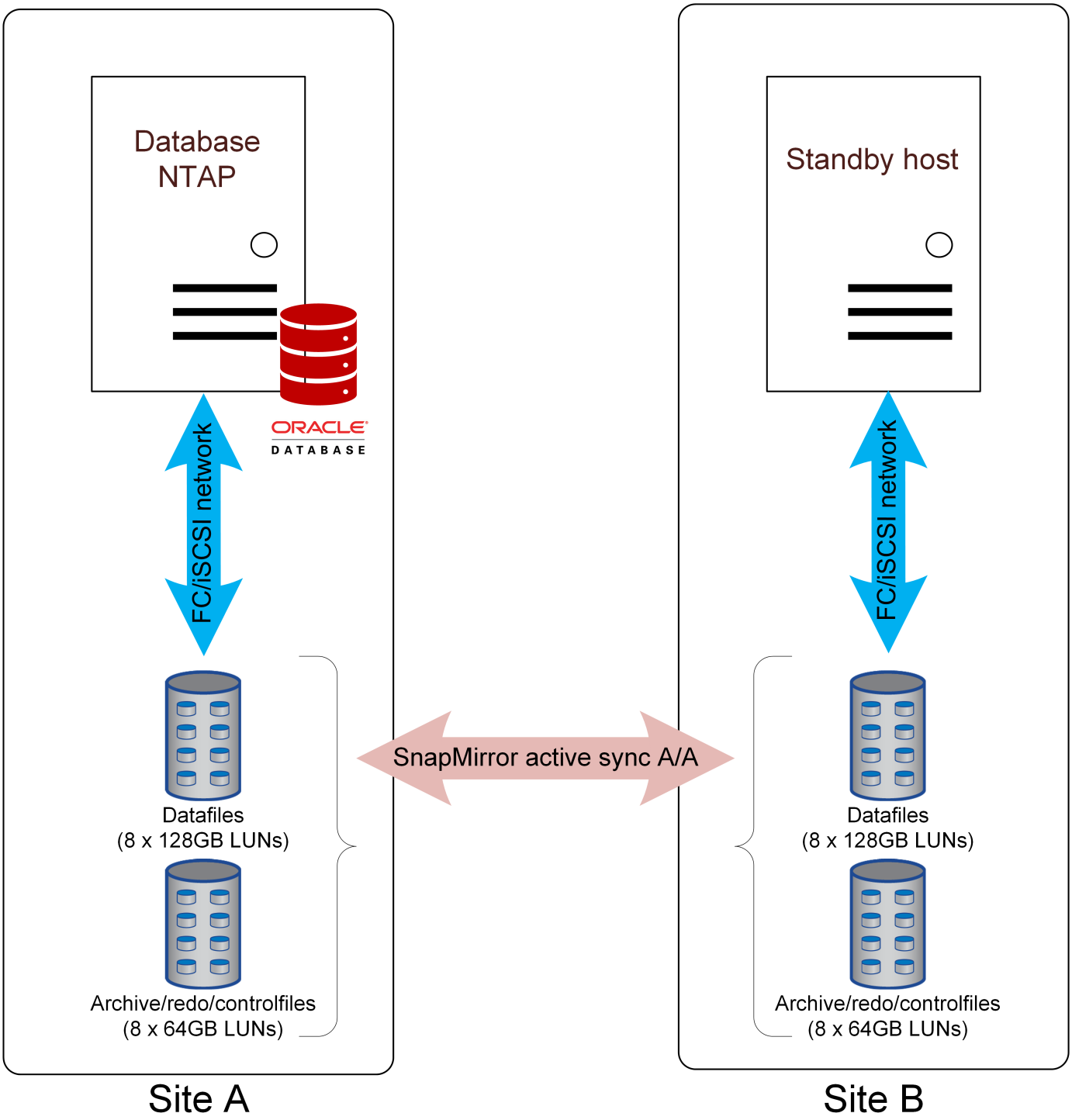
La meilleure architecture dépend des besoins de l’entreprise. Par exemple, si l’objectif est de bénéficier d’une protection RPO=0 contre la perte de données, mais que l’objectif RTO est assoupli, l’utilisation de bases de données Oracle Single instance et la réplication des LUN avec SM-AS peuvent suffire et être moins coûteuses d’un standard de licences Oracle. Toute panne du site distant n’interrompt pas les opérations, et la perte du site principal entraînerait la présence de LUN en ligne et prêtes à être utilisées sur le site survivant.

Si le RTO était plus strict, l’automatisation actif-passif de base via des scripts ou des clusters comme Pacemaker ou Ansible améliorerait le délai de basculement. Par exemple, VMware HA peut être configuré pour détecter une panne de VM sur le site principal et activer cette dernière sur le site distant.

Enfin, pour un basculement extrêmement rapide, Oracle RAC peut être déployé sur plusieurs sites. L’objectif de délai de restauration serait essentiellement égal à zéro, car la base de données serait en ligne et disponible à tout moment sur les deux sites.

Instance unique Oracle

Les exemples décrits ci-dessous illustrent certaines des nombreuses options de déploiement des bases de données Oracle Single instance avec la réplication SnapMirror Active Sync.



Basculement avec un système d'exploitation préconfiguré

La synchronisation active SnapMirror fournit une copie synchrone des données au niveau du site de reprise d'activité. Toutefois, la mise à disposition des données requiert un système d'exploitation et les applications associées. L'automatisation de base peut considérablement améliorer le délai de basculement de l'environnement global. Les produits Clusterware tels que Pacemaker sont souvent utilisés pour créer un cluster sur les sites et, dans la plupart des cas, le processus de basculement peut être piloté par des scripts simples.

En cas de perte des nœuds principaux, le cluster (ou les scripts) mettra les bases de données en ligne sur le site secondaire. Une option consiste à créer des serveurs de secours préconfigurés pour les ressources SAN

qui constituent la base de données. En cas de défaillance du site principal, le logiciel de mise en cluster ou l'alternative scriptée effectue une séquence d'actions similaires à celles décrites ci-dessous :

1. Détection d'une défaillance du site principal
2. Effectuez la détection des LUN FC ou iSCSI
3. Montage de systèmes de fichiers et/ou montage de groupes de disques ASM
4. Démarrage de la base de données

Cette approche doit avant tout se passer d'un système d'exploitation en cours d'exécution sur le site distant. Elles doivent être préconfigurées avec des binaires Oracle, ce qui signifie également que des tâches telles que l'application de correctifs Oracle doivent être effectuées sur les sites principal et de secours. Les binaires Oracle peuvent également être mis en miroir vers le site distant et montés en cas d'incident.

La procédure d'activation réelle est simple. Les commandes telles que la découverte de LUN ne nécessitent que quelques commandes par port FC. Le montage du système de fichiers n'est rien de plus qu'une `mount` commande et les bases de données et ASM peuvent être démarrés et arrêtés sur l'interface de ligne de commande à l'aide d'une seule commande.

Basculement avec un système d'exploitation virtualisé

Le basculement des environnements de base de données peut être étendu pour inclure le système d'exploitation lui-même. En théorie, ce basculement peut être effectué avec des LUN de démarrage, mais le plus souvent avec un système d'exploitation virtualisé. La procédure est similaire aux étapes suivantes :

1. Détection d'une défaillance du site principal
2. Montage des datastores hébergeant les machines virtuelles du serveur de base de données
3. Démarrage des machines virtuelles
4. Démarrage manuel des bases de données ou configuration des machines virtuelles pour démarrer automatiquement les bases de données.

Par exemple, un cluster ESX peut couvrir des sites. En cas d'incident, les machines virtuelles peuvent être mises en ligne sur le site de reprise après incident après le basculement.

Protection contre les défaillances du stockage

Le diagramme ci-dessus montre l'utilisation de "accès non uniforme", où le SAN n'est pas étendu entre les sites. Cela peut être plus simple à configurer et, dans certains cas, peut être la seule option étant donné les fonctionnalités SAN actuelles, mais cela signifie également que la défaillance du système de stockage principal entraînerait une panne de la base de données jusqu'à ce que l'application ait été ratée.

Pour une résilience supplémentaire, la solution pourrait être déployée avec "accès uniforme". Cela permettrait aux applications de continuer à fonctionner en utilisant les chemins annoncés à partir du site opposé.

RAC étendu Oracle

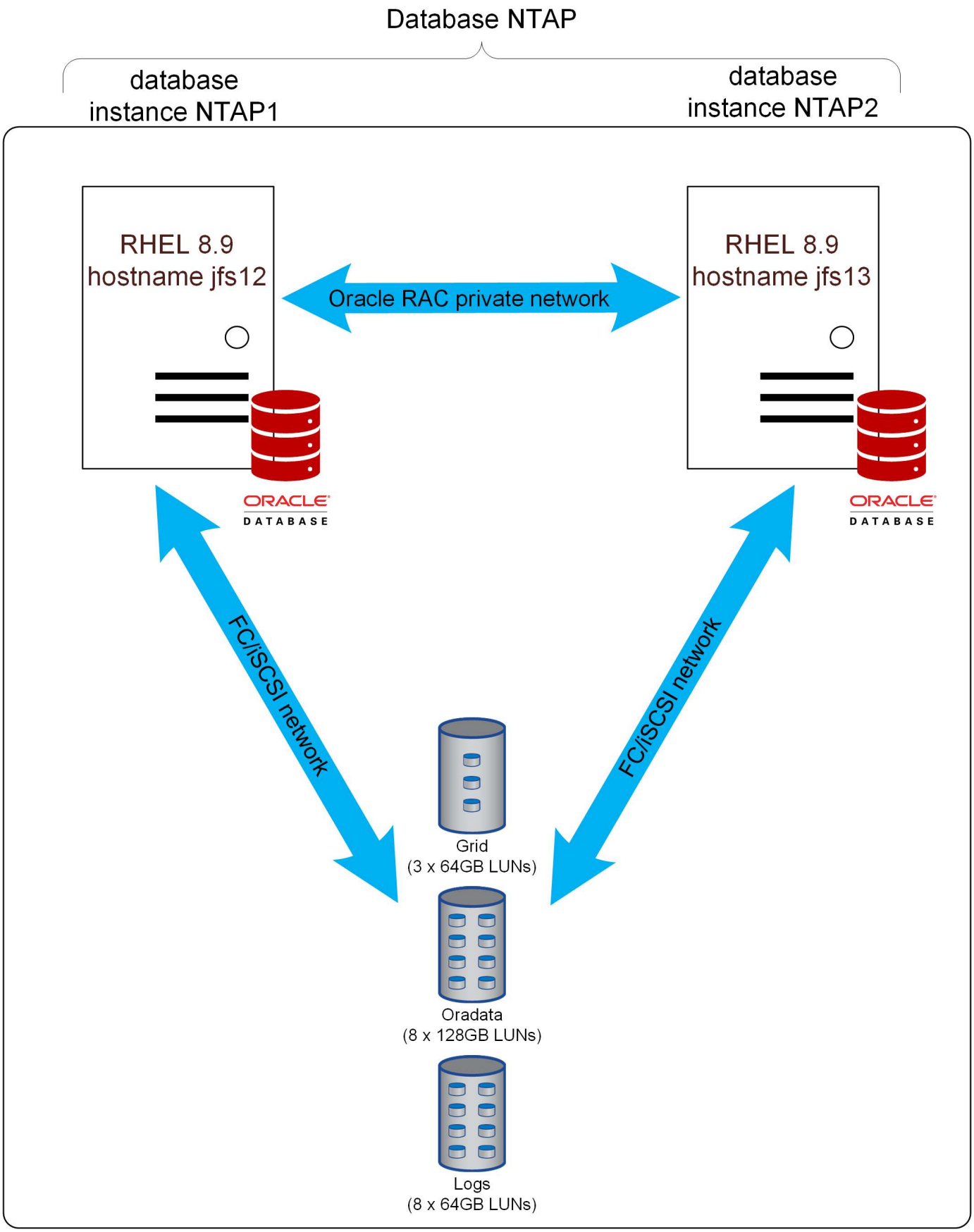
De nombreux clients optimisent leur RTO en étendant un cluster Oracle RAC sur plusieurs sites, offrant une configuration entièrement active/active. La conception globale devient plus complexe car elle doit inclure la gestion du quorum d'Oracle RAC.

La mise en cluster RAC étendue traditionnelle s'est appuyée sur la mise en miroir ASM pour assurer la protection des données. Cette approche fonctionne, mais elle implique également de nombreuses étapes manuelles de configuration et entraîne une surcharge de l'infrastructure réseau. À l'inverse, la réplication des

données peut être prise en charge par la synchronisation active SnapMirror, ce qui simplifie considérablement la solution. Les opérations telles que la synchronisation, la resynchronisation après les interruptions, les basculements et la gestion du quorum sont plus simples. En outre, le SAN n’a pas besoin d’être distribué entre les sites, ce qui simplifie la conception et la gestion du SAN.

La réplication

Pour comprendre la fonctionnalité RAC sur SnapMirror Active Sync, il est essentiel de considérer le stockage comme un ensemble unique de LUN hébergés sur un stockage en miroir. Par exemple :



Il n'y a pas de copie principale ou miroir. Pour schématiser, il n'y a qu'une seule copie de chaque LUN et cette LUN est disponible sur les chemins SAN situés sur deux systèmes de stockage différents. Du point de vue de l'hôte, il n'y a pas de basculement de stockage ; il y a des changements de chemin. Plusieurs défaillances

peuvent entraîner la perte de certains chemins vers la LUN, tandis que les autres chemins restent en ligne. La synchronisation active SnapMirror garantit la disponibilité des mêmes données sur tous les chemins opérationnels.

Configuration de stockage sous-jacente

Dans cet exemple de configuration, les disques ASM sont configurés de la même manière que dans n'importe quelle configuration RAC à site unique sur le stockage d'entreprise. Étant donné que le système de stockage assure la protection des données, la redondance ASM externe est utilisée.

Accès uniforme ou non informé

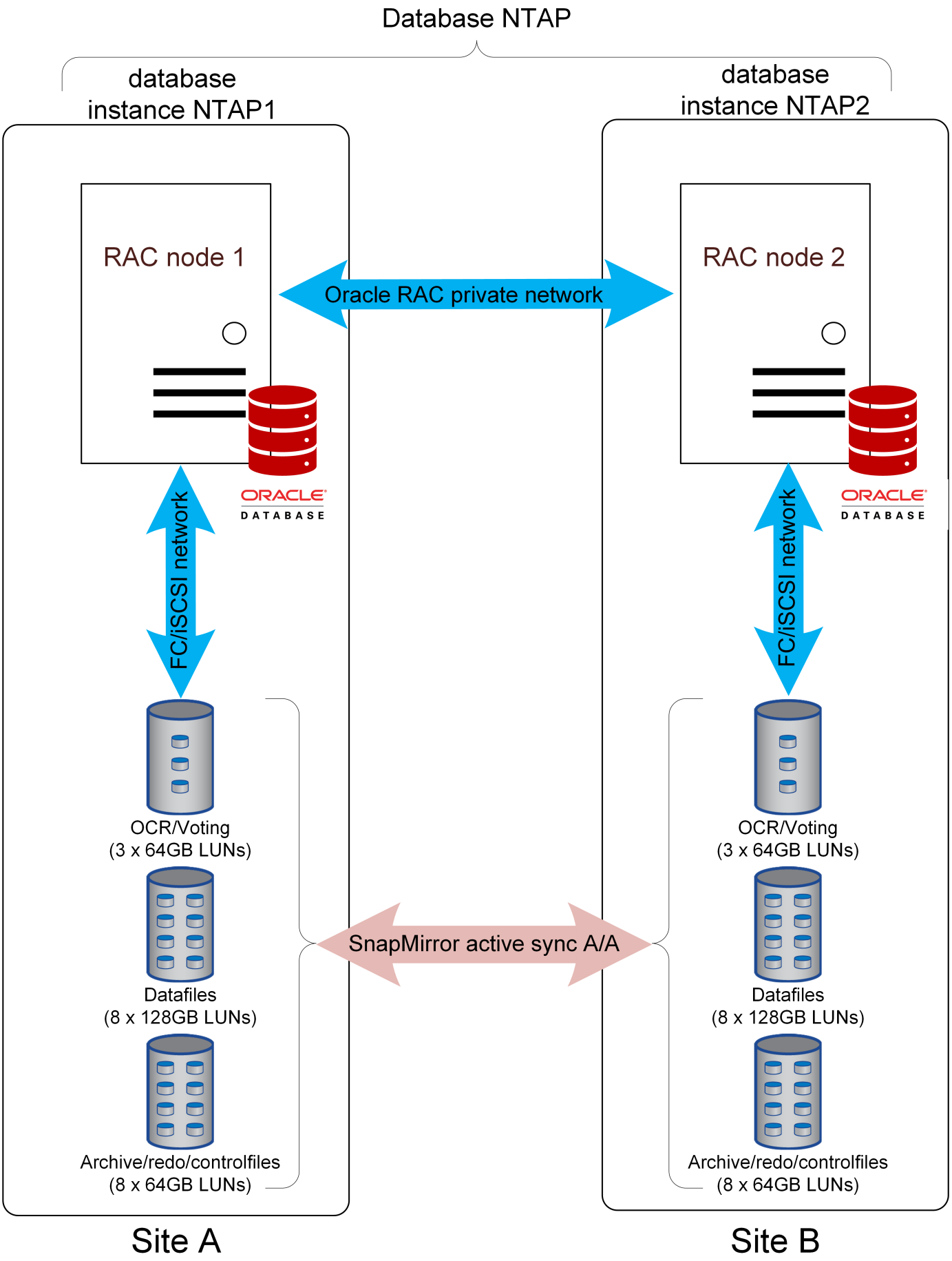
L'élément le plus important à prendre en compte avec Oracle RAC sur SnapMirror Active Sync est de savoir s'il faut utiliser un accès uniforme ou non.

Un accès uniforme signifie que chaque hôte peut voir les chemins sur les deux clusters. L'accès non uniforme signifie que les hôtes peuvent uniquement voir les chemins vers le cluster local.

Aucune de ces options n'est spécifiquement recommandée ou déconseillée. Certains clients ont facilement accès à la fibre noire pour connecter les sites, d'autres ne disposent pas d'une telle connectivité ou leur infrastructure SAN ne prend pas en charge l'ISL longue distance.

Accès non uniforme

L'accès non uniforme est plus simple à configurer du point de vue du SAN.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
ORACLE DATABASE
ORACLE DATABASE
FC/iSCSI network
FC/iSCSI network
OCR/Voting (3 x 64GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
SnapMirror active sync A/A
OCR/Voting (3 x 64GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Site A
Site B

L'inconvénient principal de cette "accès non uniforme" approche est que la perte de la connectivité ONTAP site à site ou la perte d'un système de stockage entraînera la perte des instances de base de données sur un site. Cela n'est évidemment pas souhaitable, mais cela peut constituer un risque acceptable en échange d'une configuration SAN plus simple.

Accès uniforme

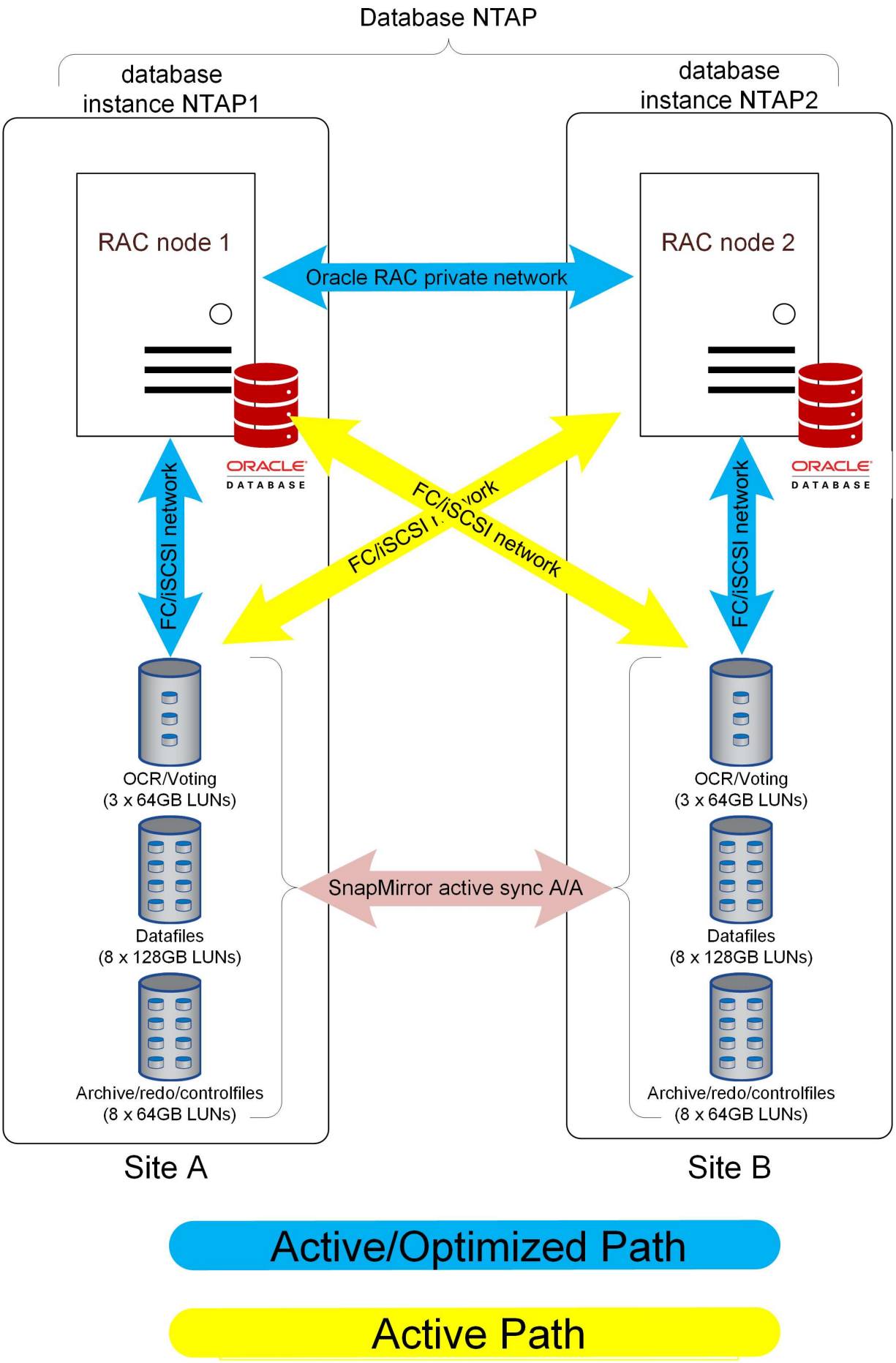
L'accès uniforme requiert l'extension du SAN sur les sites. Le principal avantage est que la perte d'un système de stockage n'entraîne pas la perte d'une instance de base de données. Au lieu de cela, cela entraînerait une modification des chemins d'accès multiples dans lesquels les chemins sont actuellement utilisés.

Il existe plusieurs façons de configurer l'accès non uniforme.

> Dans les schémas ci-dessous, il existe également des chemins actifs mais non optimisés qui seraient utilisés en cas de défaillances simples du contrôleur, mais ces chemins ne sont pas affichés dans l'intérêt de simplifier les diagrammes.

AFF avec paramètres de proximité

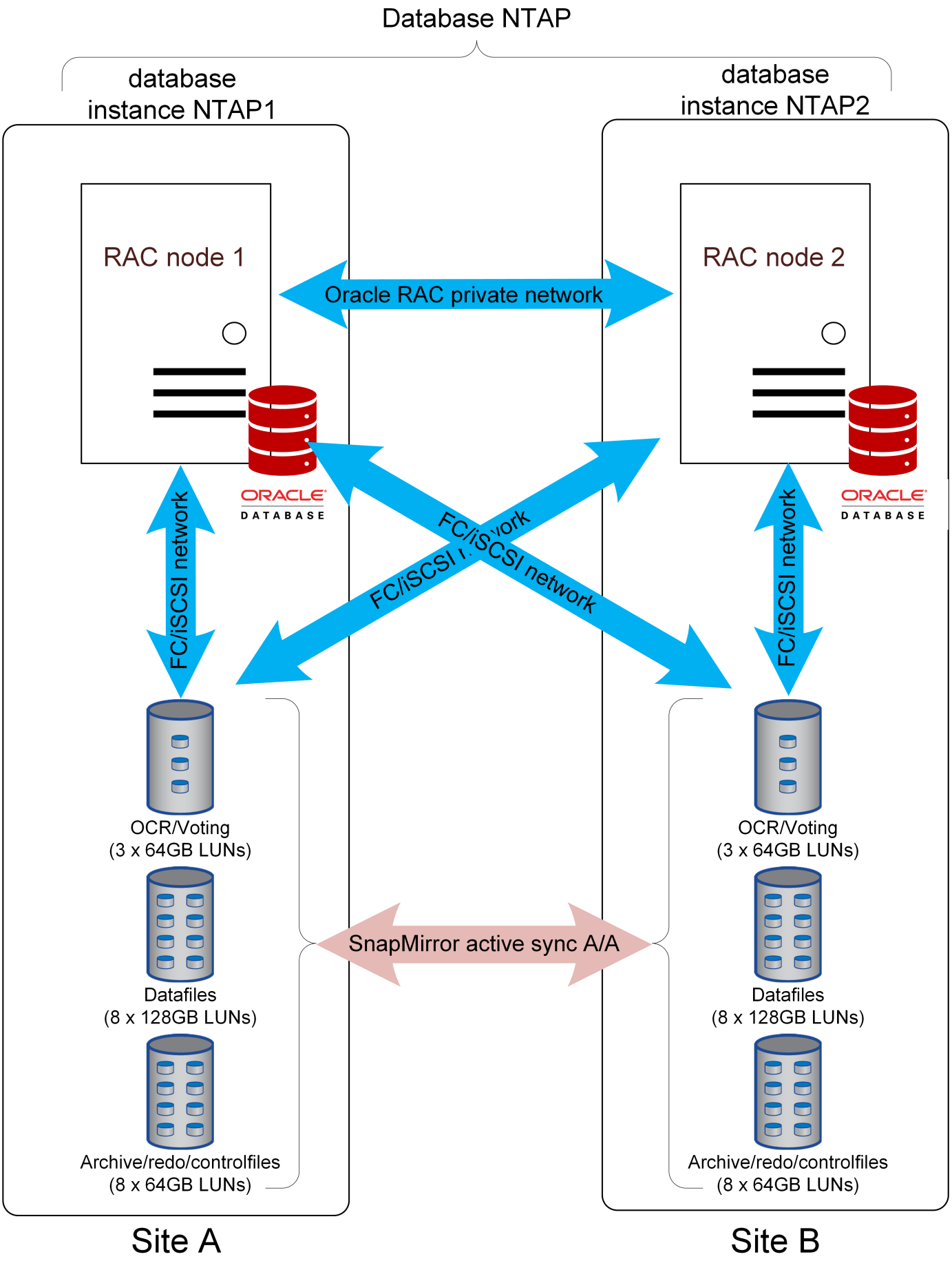
En cas de latence importante entre les sites, les systèmes AFF peuvent être configurés avec des paramètres de proximité des hôtes. Cela permet à chaque système de stockage d'identifier les hôtes locaux et distants, et d'attribuer les priorités de chemin en conséquence.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
ORACLE DATABASE
ORACLE DATABASE
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
OCR/Voting (3 x 64GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
SnapMirror active sync A/A
OCR/Voting (3 x 64GB LUNs)
Datafiles (8 x 128GB LUNs)
Archive/redo/controlfiles (8 x 64GB LUNs)
Site A
Site B
Active/Optimized Path
Active Path

En fonctionnement normal, chaque instance Oracle utilisera de préférence les chemins locaux actifs/optimisés. Par conséquent, toutes les lectures seront traitées par la copie locale des blocs. La latence est ainsi la plus faible possible. Les E/S d'écriture sont envoyées de la même manière vers le contrôleur local. L'E/S doit toujours être répliquée avant d'être reconnue, ce qui entraîne toujours une latence supplémentaire en traversant le réseau site à site, mais cela ne peut pas être évité dans une solution de réplication synchrone.

ASA / AFF sans paramètres de proximité

S'il n'y a pas de latence significative entre les sites, les systèmes AFF peuvent être configurés sans paramètres de proximité des hôtes, ou ASA peut être utilisé.


Database NTAP
database instance NTAP1
database instance NTAP2
RAC node 1
RAC node 2
Oracle RAC private network
ORACLE DATABASE
ORACLE DATABASE
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
FC/iSCSI network
OCR/Voting
(3 x 64GB LUNs)
Datafiles
(8 x 128GB LUNs)
Archive/redo/controlfiles
(8 x 64GB LUNs)
SnapMirror active sync A/A
OCR/Voting
(3 x 64GB LUNs)
Datafiles
(8 x 128GB LUNs)
Archive/redo/controlfiles
(8 x 64GB LUNs)
Site A
Site B

Chaque hôte pourra utiliser tous les chemins opérationnels sur les deux systèmes de stockage. Cela améliore considérablement les performances en permettant à chaque hôte d'exploiter le potentiel de performance de deux clusters, et non d'un seul.

Avec ASA, non seulement tous les chemins vers les deux clusters sont considérés comme actifs et optimisés, mais les chemins sur les contrôleurs partenaires sont également actifs. Il en résulte des chemins SAN entièrement actifs sur l'ensemble du cluster, à tout moment.

> Les systèmes ASA peuvent également être utilisés dans une configuration d'accès non uniforme. Étant donné qu'il n'existe aucun chemin entre les sites, les performances ne seraient pas améliorées par le franchissement de l'ISL par les E/S.

Disjoncteur d'attache RAC

Bien que le RAC étendu utilisant la synchronisation active SnapMirror soit une architecture symétrique par rapport aux E/S, il existe une exception qui est connectée à la gestion du split-brain.

Que se passe-t-il si le lien de réplication est perdu et qu'aucun des sites n'a le quorum ? Que doit-on faire ? Cette question s'applique à la fois au comportement d'Oracle RAC et de ONTAP. Si les modifications ne peuvent pas être répliquées sur tous les sites et que vous souhaitez reprendre les opérations, l'un des sites devra survivre et l'autre site devra être indisponible.

Le système "Médiateur de ONTAP" répond à cette exigence au niveau de la couche ONTAP. Il existe plusieurs options pour le tricover RAC.

Disjoncteurs Oracle

La meilleure méthode pour gérer les risques Oracle RAC split-brain consiste à utiliser un nombre impair de nœuds RAC, de préférence à l'aide d'un Tiebreaker 3rd site. Si un troisième site n'est pas disponible, l'instance Tiebreaker pourrait être placée sur un site des deux sites, ce qui la désignerait en fait un site de survivant préféré.

Oracle et CSS_Critical

Avec un nombre pair de nœuds, le comportement par défaut d'Oracle RAC est que l'un des nœuds du cluster sera considéré plus important que les autres nœuds. Le site avec ce nœud de priorité supérieure survivra à l'isolation du site tandis que les nœuds de l'autre site seront supprimés. La hiérarchisation est basée sur plusieurs facteurs, mais vous pouvez également contrôler ce comportement à l'aide du `css_critical` paramètre.

Dans l'"exemple"architecture, les noms d'hôte des nœuds RAC sont jfs12 et jfs13. Les paramètres actuels de `css_critical` sont les suivants :

```
[root@jfs12 ~]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.

[root@jfs13 trace]# /grid/bin/crsctl get server css_critical
CRS-5092: Current value of the server attribute CSS_CRITICAL is no.
```

Si vous voulez que le site avec jfs12 soit le site préféré, définissez cette valeur sur oui sur un site Un noeud et

redémarrez les services.

```
[root@jfs12 ~]# /grid/bin/crsctl set server css_critical yes
CRS-4416: Server attribute 'CSS_CRITICAL' successfully changed. Restart
Oracle High Availability Services for new value to take effect.

[root@jfs12 ~]# /grid/bin/crsctl stop crs
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'jfs12'
CRS-2673: Attempting to stop 'ora.crsd' on 'jfs12'
CRS-2790: Starting shutdown of Cluster Ready Services-managed resources on
server 'jfs12'
CRS-2673: Attempting to stop 'ora.ntap.ntappdb1.pdb' on 'jfs12'
…
CRS-2673: Attempting to stop 'ora.gipcd' on 'jfs12'
CRS-2677: Stop of 'ora.gipcd' on 'jfs12' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'jfs12' has completed
CRS-4133: Oracle High Availability Services has been stopped.

[root@jfs12 ~]# /grid/bin/crsctl start crs
CRS-4123: Oracle High Availability Services has been started.
```

Scénarios d'échec

Présentation

La planification d'une architecture complète d'applications de synchronisation active SnapMirror nécessite de comprendre comment les SM-AS répondront dans divers scénarios de basculement planifiés et non planifiés.

Pour les exemples suivants, supposons que le site A est configuré comme le site préféré.

Perte de la connectivité de réplication

Si la réplication SM-AS est interrompue, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A (site préféré)

Le résultat de l'échec de la liaison de réplication sur le site préféré sera une pause d'environ 15 secondes dans le traitement des E/S d'écriture, car ONTAP relance les opérations d'écriture répliquées avant de déterminer que la liaison de réplication est véritablement inaccessible. Au bout de 15 secondes, le site A du système reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Panne du système de stockage

Le résultat d'une défaillance du système de stockage est presque identique au résultat de la perte du lien de réplication. Le site survivant devrait subir une pause d'E/S d'environ 15 seconde. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Perte du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ». En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

La raison pour laquelle un médiateur est nécessaire pour un basculement automatisé sécurisé est parce qu'il serait autrement impossible à un cluster de stockage de déterminer si la perte de la communication bidirectionnelle était le résultat d'une panne du réseau ou d'une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.
- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, y compris la défaillance du site ou du système de stockage non privilégié, le site préféré libère l'état RPO=0 et reprend le traitement des E/S de lecture et d'écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration des services

Après résolution d'une panne, par exemple lors de la restauration de la connectivité site à site ou de la mise sous tension d'un système défaillant, les terminaux de synchronisation active SnapMirror détectent automatiquement la présence d'une relation de réplication défectueuse et la raverront à l'état RPO=0. Une fois la réplication synchrone rétablie, les chemins défaillants se reconnectent.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement. Cela dépend de l'application et de la façon dont elle est configurée et, en général, de telles tâches peuvent être facilement automatisées. La fonctionnalité ONTAP elle-même est dotée d'une fonctionnalité d'autorétablissement et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0.

Basculement manuel

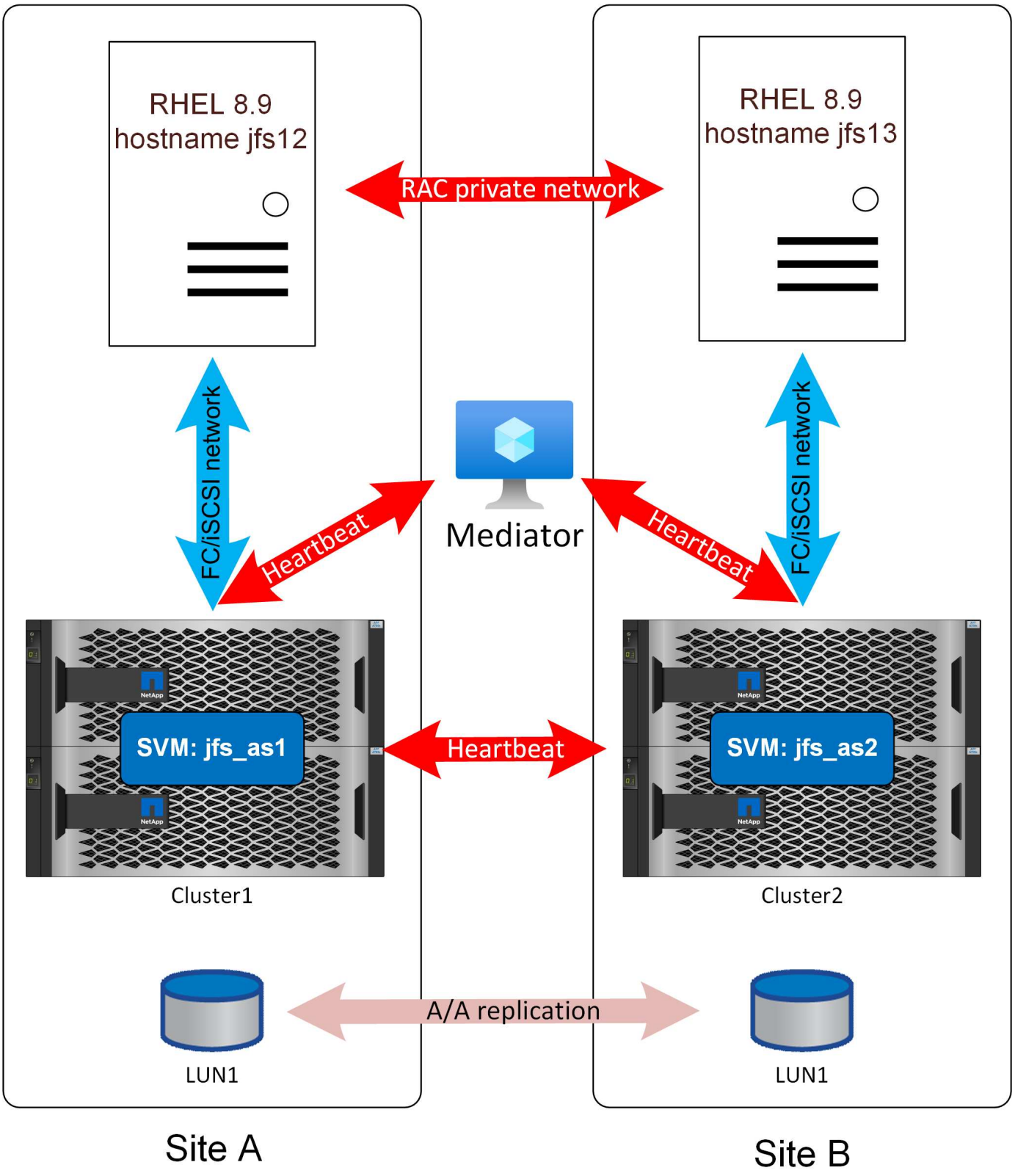
La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

Exemple d'architecture

Les exemples détaillés de défaillances présentés dans cette section sont basés sur l'architecture présentée ci-dessous.

> Il ne s'agit que de l'une des nombreuses options pour les bases de données Oracle sur la synchronisation active SnapMirror. Cette conception a été choisie parce qu'elle illustre certains des scénarios les plus complexes.

Dans cette conception, supposons que le site A est défini sur "site préféré".



Échec de l'interconnexion du RAC

La perte du lien de réplication RAC Oracle produira un résultat similaire à la perte de la connectivité SnapMirror, sauf que les délais d'expiration seront plus courts par défaut. Dans les paramètres par défaut, un nœud RAC Oracle attend 200 secondes après une

perte de connectivité du stockage avant d'être supprimé, mais il n'attend que 30 secondes après la perte du signal de détection du réseau RAC.

Les messages CRS sont similaires à ceux indiqués ci-dessous. Vous pouvez voir le délai d'expiration de 30 secondes. Comme css_Critical a été défini sur jfs12, situé sur le site A, ce sera le site pour survivre et jfs13 sur le site B sera supprimé.

```
2024-09-12 10:56:44.047 [ONMD(3528)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 6.980 seconds
2024-09-12 10:56:48.048 [ONMD(3528)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.980 seconds
2024-09-12 10:56:51.031 [ONMD(3528)]CRS-1607: Node jfs13 is being evicted
in cluster incarnation 621599354; details at (:CSSNM00007:) in
/gridbase/diag/crs/jfs12/crs/trace/onmd.trc.
2024-09-12 10:56:52.390 [CRSD(6668)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:33194;', interface list of remote node 'jfs13' is
'192.168.30.2:33621;'.
2024-09-12 10:56:55.683 [ONMD(3528)]CRS-1601: CSSD Reconfiguration
complete. Active nodes are jfs12 .
2024-09-12 10:56:55.722 [CRSD(6668)]CRS-5504: Node down event reported for
node 'jfs13'.
2024-09-12 10:56:57.222 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'Generic'.
2024-09-12 10:56:57.224 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'ora.NTAP'.
```

Échec de communication SnapMirror

Si la liaison de réplication SnapMirror active Sync, l'E/S d'écriture ne peut pas être terminée, car un cluster ne peut pas répliquer les modifications sur le site opposé.

Site A

Le site A qui présente une défaillance de liaison de réplication entraînera une pause d'environ 15 secondes dans le traitement des E/S d'écriture au fur et à mesure que ONTAP tente de répliquer des écritures avant de déterminer que la liaison de réplication est réellement inutilisable. Au bout de 15 secondes, le cluster ONTAP sur le site A reprend le traitement des E/S de lecture et d'écriture. Les chemins SAN ne changent pas et les LUN restent en ligne.

Site B

Le site B n'étant pas le site privilégié de synchronisation active SnapMirror, ses chemins de LUN deviennent indisponibles au bout de 15 secondes environ.

Le lien de réplication a été coupé à l'horodatage 15:19:44. Le premier avertissement d'Oracle RAC arrive 100 secondes plus tard lorsque le délai d'expiration de 200 secondes (contrôlé par le paramètre Oracle RAC disktimeout) approche.

```
2024-09-10 15:21:24.702 [ONMD(2792)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99340 milliseconds.
2024-09-10 15:22:14.706 [ONMD(2792)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49330 milliseconds.
2024-09-10 15:22:44.708 [ONMD(2792)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19330 milliseconds.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.716 [ONMD(2792)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.731 [OCSSD(2794)]CRS-1652: Starting clean up of CRSD
resources.
```

Une fois que le délai d'expiration du disque de vote de 200 secondes a été atteint, ce nœud RAC Oracle s'expulse automatiquement du cluster et redémarre.

Échec total de l'interconnectivité réseau

Si la liaison de réplication entre les sites est totalement perdue, la synchronisation active SnapMirror et la connectivité RAC Oracle seront interrompues.

La détection d'Oracle RAC à cerveau divisé dépend du pulsation du stockage Oracle RAC. Si la perte de la connectivité site à site entraîne la perte simultanée du signal de détection du réseau RAC et des services de réplication du stockage, les sites RAC ne pourront pas communiquer entre sites via l'interconnexion RAC ou les disques de vote RAC. Le résultat d'un ensemble de nœuds à numéro pair peut être l'exclusion des deux sites sous les paramètres par défaut. Le comportement exact dépend de la séquence des événements et de la synchronisation des sondages de pulsation du réseau RAC et du disque.

Le risque d'une panne sur deux sites peut être résolu de deux manières. Tout d'abord, une "disjoncteur d'attache" configuration peut être utilisée.

Si aucun site tiers n'est disponible, ce risque peut être résolu en ajustant le paramètre misscount sur le cluster RAC. Sous les valeurs par défaut, le délai d'expiration de la pulsation réseau du RAC est de 30 secondes. Il est généralement utilisé par RAC pour identifier les nœuds RAC défaillants et les supprimer du cluster. Il dispose également d'une connexion à la pulsation du disque de vote.

Si, par exemple, le conduit transportant le trafic intersite pour Oracle RAC et les services de réplication de stockage est coupé par une pelle rétro, le compte à rebours des erreurs de 30 secondes commence. Si le nœud du site RAC préféré ne peut pas rétablir le contact avec le site opposé dans les 30 secondes et qu'il ne peut pas utiliser les disques de vote pour confirmer que le site opposé est en panne dans la même fenêtre de 30 secondes, les nœuds du site préféré seront également supprimés. Il en résulte une interruption complète de la base de données.

Selon le moment où l'interrogation du compte erroné se produit, 30 secondes peuvent ne pas suffire à la temporisation de la synchronisation active SnapMirror et à permettre au stockage du site préféré de reprendre les services avant l'expiration de la fenêtre de 30 secondes. Cette fenêtre de 30 secondes peut être augmentée.

```
[root@jfs12 ~]# /grid/bin/crsctl set css misscount 100
CRS-4684: Successful set of parameter misscount to 100 for Cluster
Synchronization Services.
```

Cette valeur permet au système de stockage sur le site préféré de reprendre les opérations avant que le délai d'erreur n'expire. Le résultat sera alors la suppression uniquement des nœuds sur le site où les chemins de LUN ont été supprimés. Exemple ci-dessous :

```
2024-09-12 09:50:59.352 [ONMD(681360)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 49.570 seconds
2024-09-12 09:51:10.082 [CRSD(682669)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:46039;', interface list of remote node 'jfs13' is
'192.168.30.2:42037;'.
2024-09-12 09:51:24.356 [ONMD(681360)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 24.560 seconds
2024-09-12 09:51:39.359 [ONMD(681360)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 9.560 seconds
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8011: reboot advisory message
from host: jfs13, component: cssagent, with time stamp: L-2024-09-12-
09:51:47.451
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8013: reboot advisory message
text: oracssdagent is about to reboot this node due to unknown reason as
it did not receive local heartbeats for 10470 ms amount of time
2024-09-12 09:51:48.925 [ONMD(681360)]CRS-1632: Node jfs13 is being
removed from the cluster in cluster incarnation 621596607
```

Le support Oracle déconseille fortement de modifier les paramètres misscount ou disktimeout pour résoudre les problèmes de configuration. Toutefois, la modification de ces paramètres peut s'avérer justifiée et inévitable dans de nombreux cas, notamment dans les configurations de démarrage SAN, de virtualisation et de réplication du stockage. Si, par exemple, vous avez rencontré des problèmes de stabilité avec un réseau SAN

ou IP qui ont entraîné des expulsions RAC, vous devez résoudre le problème sous-jacent et ne pas facturer les valeurs de l'erreur de décompte ou du dépassement de disque. La modification des délais pour résoudre les erreurs de configuration masque un problème et non pas résout un problème. La modification de ces paramètres pour configurer correctement un environnement RAC basé sur les aspects de conception de l'infrastructure sous-jacente est différente et est conforme aux instructions de support Oracle. Avec le démarrage SAN, il est courant d'ajuster misscount jusqu'à 200 pour correspondre au disktimeout. Voir "ce lien" pour plus d'informations.

Panne du site

Le résultat d'une défaillance du site ou du système de stockage est presque identique au résultat de la perte du lien de réplication. Le site survivant doit subir une pause d'E/S d'environ 15 secondes sur les écritures. Une fois cette période de 15 secondes écoulée, l'E/S reprend sur ce site comme d'habitude.

Si seul le système de stockage a été affecté, le nœud Oracle RAC sur le site en panne perdra les services de stockage et entrera le même compte à rebours de 200 secondes avant la suppression et le redémarrage suivant.

```
2024-09-11 13:44:38.613 [ONMD(3629)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99750 milliseconds.
2024-09-11 13:44:51.202 [ORAAGENT(5437)]CRS-5011: Check of resource "NTAP"
failed: details at "(:CLSN00007:)" in
"/gridbase/diag/crs/jfs13/crs/trace/crsd_oraagent_oracle.trc"
2024-09-11 13:44:51.798 [ORAAGENT(75914)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 75914
2024-09-11 13:45:28.626 [ONMD(3629)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49730 milliseconds.
2024-09-11 13:45:33.339 [ORAAGENT(76328)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 76328
2024-09-11 13:45:58.629 [ONMD(3629)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19730 milliseconds.
2024-09-11 13:46:18.630 [ONMD(3629)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-11 13:46:18.631 [ONMD(3629)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.638 [ONMD(3629)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.651 [OCSSD(3631)]CRS-1652: Starting clean up of CRSD
resources.
```

L'état du chemin SAN sur le nœud RAC qui a perdu des services de stockage se présente comme suit :

```
oradata7 (3600a0980383041334a3f55676c697347) dm-20 NETAPP,LUN C-Mode
size=128G features='3 queue_if_no_path pg_init_retries 50' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
| `- 34:0:0:18 sdam 66:96  failed faulty running
`-+- policy='service-time 0' prio=0 status=enabled
  `- 33:0:0:18 sdaj 66:48  failed faulty running
```

L'hôte linux a détecté la perte des chemins beaucoup plus rapidement que 200 secondes, mais du point de vue de la base de données, les connexions client à l'hôte sur le site défaillant seront toujours bloquées pendant 200 secondes sous les paramètres Oracle RAC par défaut. Les opérations complètes de la base de données ne reprendront qu'une fois la suppression terminée.

Pendant ce temps, le nœud Oracle RAC sur le site opposé enregistre la perte de l'autre nœud RAC. Dans le cas contraire, le système continue de fonctionner normalement.

```
2024-09-11 13:46:34.152 [ONMD(3547)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 14.020 seconds
2024-09-11 13:46:41.154 [ONMD(3547)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 7.010 seconds
2024-09-11 13:46:46.155 [ONMD(3547)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.010 seconds
2024-09-11 13:46:46.470 [OHASD(1705)]CRS-8011: reboot advisory message
from host: jfs13, component: cssmonit, with time stamp: L-2024-09-11-
13:46:46.404
2024-09-11 13:46:46.471 [OHASD(1705)]CRS-8013: reboot advisory message
text: At this point node has lost voting file majority access and
oracssdmonitor is rebooting the node due to unknown reason as it did not
receive local hearbeats for 28180 ms amount of time
2024-09-11 13:46:48.173 [ONMD(3547)]CRS-1632: Node jfs13 is being removed
from the cluster in cluster incarnation 621516934
```

Défaillance du médiateur

Le service médiateur ne contrôle pas directement les opérations de stockage. Il fonctionne comme un chemin de contrôle alternatif entre les clusters. Il existe principalement pour automatiser le basculement sans les risques associés à un scénario « split-brain ».

En conditions normales de fonctionnement, chaque cluster réplique les modifications apportées à son partenaire et chaque cluster peut donc vérifier que le cluster partenaire est en ligne et qu'il transmet les données. Si le lien de réplication échoue, la réplication s'arrête.

Un médiateur est nécessaire pour des opérations automatisées sécurisées, car il serait autrement impossible pour les clusters de stockage de déterminer si la perte de la communication bidirectionnelle était due à une panne du réseau ou à une défaillance réelle du stockage.

Le médiateur fournit un chemin alternatif pour chaque cluster afin de vérifier l'état de santé de son partenaire. Les scénarios sont les suivants :

- Si un cluster peut contacter directement son partenaire, les services de réplication sont opérationnels. Aucune action requise.
- Si un site privilégié ne peut pas contacter son partenaire directement ou via le médiateur, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site préféré va ensuite publier l'état RPO=0 et continuer à traiter les E/S en lecture et en écriture.
- Si un site non préféré ne peut pas contacter directement son partenaire, mais peut le contacter via le médiateur, il mettra ses chemins hors ligne et attend le retour de la connexion de réplication.

- Si un site non privilégié ne peut pas contacter son partenaire directement ou via un médiateur opérationnel, il suppose que le partenaire est réellement indisponible ou a été isolé et a mis ses chemins LUN hors ligne. Le site non privilégié va ensuite publier l'état RPO=0 et continuer le traitement des E/S en lecture et en écriture. Il assumera le rôle de la source de réplication et deviendra le nouveau site préféré.

Si le médiateur n'est pas disponible :

- En cas de défaillance des services de réplication, quelle qu'en soit la raison, le site préféré libère l'état RPO=0 et reprend le traitement des E/S en lecture et en écriture. Le site non préféré mettra ses chemins hors ligne.
- La défaillance du site préféré entraînera une panne, car le site non préféré ne pourra pas vérifier que le site opposé est réellement hors ligne et, par conséquent, il ne serait pas sûr que le site non préféré puisse reprendre ses services.

Restauration du service

SnapMirror propose une fonctionnalité d'autorétablissement. La synchronisation active SnapMirror détecte automatiquement la présence d'une relation de réplication défectueuse et la ramène à un état RPO=0. Une fois la réplication synchrone rétablie, les chemins reviennent en ligne.

Dans de nombreux cas, les applications en cluster détectent automatiquement le retour des chemins défaillants, et ces applications sont également reconnectées. Dans d'autres cas, une analyse SAN au niveau de l'hôte peut être nécessaire ou les applications doivent être reconnectées manuellement.

Cela dépend de l'application et de la façon dont elle est configurée et, en général, ces tâches peuvent être facilement automatisées. La synchronisation active SnapMirror elle-même est auto-fixing et ne nécessite aucune intervention de l'utilisateur pour reprendre les opérations de stockage avec un objectif de point de récupération de 0 une fois l'alimentation et la connectivité restaurées.

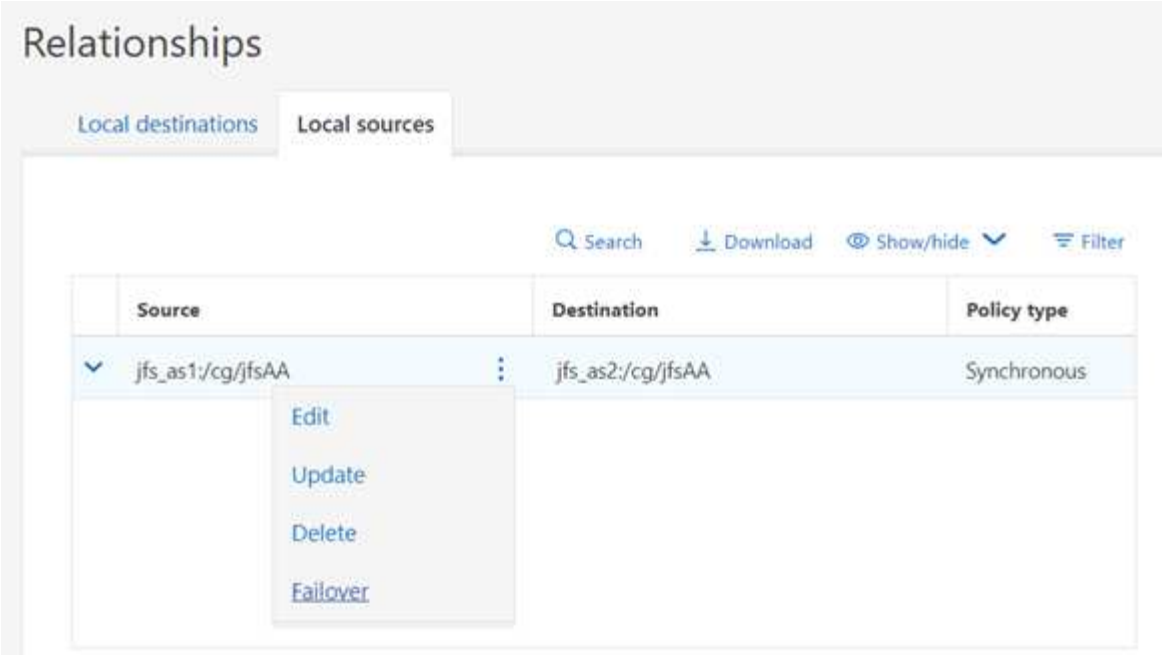
Basculement manuel

Le terme « basculement » ne fait pas référence au sens de la réplication avec la synchronisation active SnapMirror, car il s'agit d'une technologie de réplication bidirectionnelle. En revanche, le terme « basculement » désigne le système de stockage qui sera le site privilégié en cas de défaillance.

Par exemple, vous pouvez effectuer un basculement pour modifier le site préféré avant d'arrêter un site pour des raisons de maintenance ou avant d'effectuer un test de reprise après incident.

La modification du site préféré nécessite une opération simple. L'E/S s'interrompt pendant une ou deux secondes car l'autorité sur le comportement de réplication change entre les clusters, mais l'E/S n'est pas affectée.

Exemple d'interface graphique :



Exemple de modification via l’interface de ligne de commande :

```
Cluster2::> snapmirror failover start -destination-path jfs_as2:/cg/jfsAA
[Job 9575] Job is queued: SnapMirror failover for destination
"jfs_as2:/cg/jfsAA                      ".

Cluster2::> snapmirror failover show

Source    Destination                                           Error
Path      Path        Type     Status    start-time end-time   Reason
-------- -----------  -------- --------- ---------- ---------- ----------
jfs_as1:/cg/jfsAA
         jfs_as2:/cg/jfsAA
                      planned  completed 9/11/2024  9/11/2024
                                         09:29:22   09:29:32

The new destination path can be verified as follows:


Cluster1::> snapmirror show -destination-path jfs_as1:/cg/jfsAA

                           Source Path: jfs_as2:/cg/jfsAA
                        Destination Path: jfs_as1:/cg/jfsAA
                      Relationship Type: XDP
                 Relationship Group Type: consistencygroup
                  SnapMirror Policy Type: automated-failover-duplex
                      SnapMirror Policy: AutomatedFailOverDuplex
                            Tries Limit: -
                           Mirror State: Snapmirrored
                    Relationship Status: InSync
```

Migration de la base de données Oracle

Présentation

L'exploitation des capacités d'une nouvelle plateforme de stockage implique une seule nécessité : les données doivent être placées sur le nouveau système de stockage. ONTAP simplifie le processus de migration, notamment les migrations et les mises à niveau de ONTAP vers ONTAP, les importations de LUN étrangères et les procédures d'utilisation directe du système d'exploitation hôte ou du logiciel de base de données Oracle.

> Cette documentation remplace le rapport technique *TR-4534 : migration des bases de données Oracle vers des systèmes de stockage NetApp*

Dans le cas d'un nouveau projet de base de données, cela ne pose pas de problème car les environnements de base de données et d'application sont construits en place. Cependant, la migration pose des défis

particuliers en ce qui concerne les interruptions d'activité, le temps nécessaire à la réalisation de la migration, les compétences requises et la réduction des risques.

Scripts

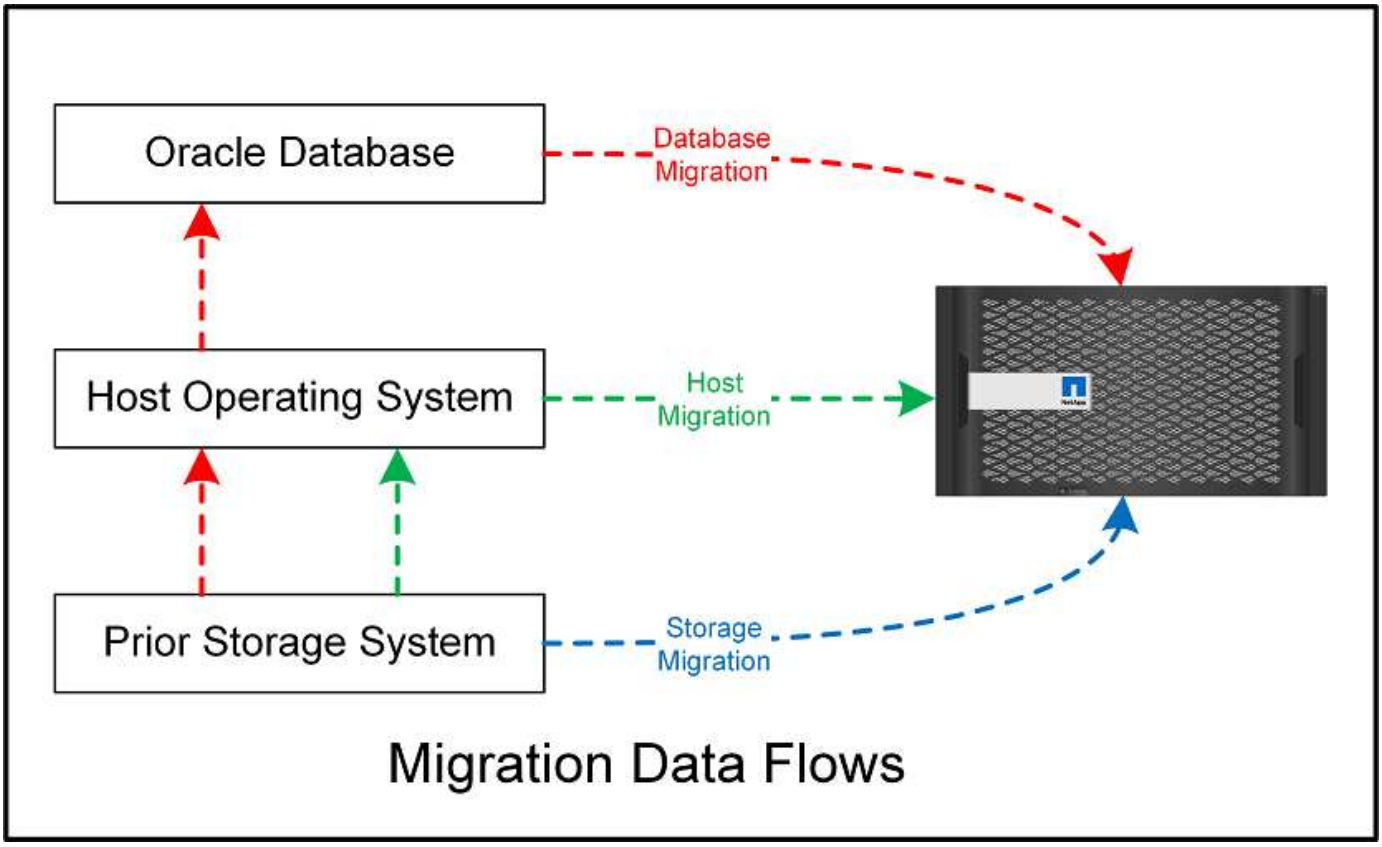
Des exemples de scripts sont fournis dans cette documentation. Ces scripts fournissent des exemples de méthodes d'automatisation de divers aspects de la migration afin de réduire le risque d'erreurs des utilisateurs. Les scripts réduisent les demandes globales de l'équipe INFORMATIQUE responsable de la migration et accélèrent le processus global. Ces scripts sont issus de projets de migration réalisés par les services professionnels de NetApp et les partenaires NetApp. Des exemples de leur utilisation sont présentés dans cette documentation.

Planification des migrations

La migration des données Oracle peut se faire à l'un des trois niveaux suivants : la base de données, l'hôte ou la baie de stockage.

Les différences résident dans la capacité du composant de la solution globale à déplacer les données : la base de données, le système d'exploitation hôte ou le système de stockage.

La figure ci-dessous présente un exemple des niveaux de migration et du flux de données. Dans le cas d'une migration au niveau de la base de données, les données sont déplacées du système de stockage d'origine vers le nouvel environnement via les couches hôte et base de données. La migration au niveau de l'hôte est similaire, mais les données ne passent pas par la couche applicative et sont écrites au nouvel emplacement à l'aide de processus hôtes. Enfin, avec la migration au niveau du stockage, une baie telle qu'un système NetApp FAS est responsable du déplacement des données.



Une migration au niveau de la base de données fait généralement référence à l'utilisation de l'envoi de journaux Oracle via une base de données de secours pour effectuer une migration au niveau de la couche

Oracle. Les migrations au niveau de l'hôte s'effectuent à l'aide de la fonctionnalité native de la configuration du système d'exploitation hôte. Cette configuration inclut des opérations de copie de fichiers à l'aide de commandes telles que cp, tar et Oracle Recovery Manager (RMAN) ou à l'aide d'un gestionnaire de volumes logiques (LVM) pour déplacer les octets sous-jacents d'un système de fichiers. Oracle Automatic Storage Management (ASM) est classé comme une fonctionnalité de niveau hôte car elle s'exécute en dessous du niveau de l'application de base de données. ASM remplace le gestionnaire de volumes logiques habituel sur un hôte. Enfin, les données peuvent être migrées au niveau de la baie de stockage, ce qui signifie en dessous du niveau du système d'exploitation.

Planification

La meilleure option de migration dépend de plusieurs facteurs, notamment de l'étendue de l'environnement à migrer, de la nécessité d'éviter les temps d'indisponibilité et des efforts globaux requis pour effectuer la migration. Les bases de données volumineuses nécessitent évidemment plus de temps et d'efforts pour la migration, mais la complexité de cette migration est minimale. Les petites bases de données peuvent être migrées rapidement. Toutefois, si des milliers d'entre elles doivent être migrées, l'ampleur des efforts peut engendrer des complications. Enfin, plus la base de données est volumineuse, plus elle est susceptible d'être stratégique, ce qui entraîne la nécessité de minimiser les temps d'indisponibilité tout en préservant un chemin « back-out ».

Voici quelques-uns des éléments à prendre en compte lors de la planification d'une stratégie de migration.

Taille des données

La taille des bases de données à migrer a de toute évidence un impact sur la planification de la migration, bien que la taille n'ait pas nécessairement un impact sur le délai de mise en service. Lorsqu'une grande quantité de données doit être migrée, la principale considération est la bande passante. Les opérations de copie s'effectuent généralement via des E/S séquentielles efficaces En guise d'estimation prudente, on suppose une utilisation de 50 % de la bande passante réseau disponible pour les opérations de copie. Par exemple, un port FC de 8 Go peut en théorie transférer environ 800 Mbit/s. Si l'on suppose une utilisation de 50 %, une base de données peut être copiée à un taux d'environ 400 Mbit/s. Ainsi, une base de données de 10 To peut être copiée en sept heures environ à ce rythme.

La migration sur de longues distances nécessite généralement une approche plus créative, comme le processus d'expédition des journaux expliqué dans "Déplacement du fichier de données en ligne". Les réseaux IP longue distance disposent rarement d'une bande passante proche des vitesses LAN ou SAN. Dans un cas, NetApp a participé à la migration à distance d'une base de données de 220 To avec des taux de génération de journaux d'archivage très élevés. L'approche choisie pour le transfert de données a été l'expédition quotidienne de bandes, parce que cette méthode offrait la bande passante maximale possible.

Nombre de bases de données

Dans de nombreux cas, le problème du déplacement d'une grande quantité de données n'est pas la taille des données, mais plutôt la complexité de la configuration qui prend en charge la base de données. Savoir qu'il faut migrer 50 To de bases de données n'est pas suffisant. Il peut s'agir d'une seule base de données stratégique de 50 To, d'un ensemble de 4 000 bases de données héritées ou d'un mélange de données de production et de données hors production. Dans certains cas, une grande partie des données est constituée de clones d'une base de données source. Il n'est pas nécessaire de migrer ces clones car ils peuvent être recréés facilement, notamment lorsque la nouvelle architecture est conçue pour exploiter les volumes NetApp FlexClone.

Pour la planification de la migration, vous devez connaître le nombre de bases de données concernées et leur priorité. À mesure que le nombre de bases de données augmente, l'option de migration privilégiée tend à être plus faible et plus faible dans la pile. Par exemple, la copie d'une seule base de données peut s'effectuer facilement avec RMAN et en cas de courte panne. Il s'agit de la réplication au niveau de l'hôte.

S’il existe 50 bases de données, il peut être plus facile d’éviter de configurer une nouvelle structure de système de fichiers pour recevoir une copie RMAN et de déplacer les données à la place. Ce processus peut être effectué en tirant parti de la migration LVM basée sur l’hôte pour déplacer les données des anciennes LUN vers les nouvelles LUN. L’équipe chargée de l’administration de la base de données (DBA) est alors détransférée vers l’équipe chargée du système d’exploitation pour que les données soient migrées de manière transparente par rapport à la base de données. La configuration du système de fichiers n’est pas modifiée.

Enfin, si 500 bases de données réparties sur 200 serveurs doivent être migrées, des options basées sur le stockage, telles que la fonctionnalité ONTAP Foreign LUN Import (FLI), peuvent être utilisées pour effectuer une migration directe des LUN.

Exigences en matière d’architecture

En général, l’organisation d’un fichier de base de données doit être modifiée pour exploiter les fonctionnalités de la nouvelle baie de stockage. Toutefois, ce n’est pas toujours le cas. Par exemple, les fonctionnalités des baies 100 % Flash EF-Series se concentrent sur les performances SAN et la fiabilité SAN. Dans la plupart des cas, les bases de données peuvent être migrées vers une baie EF-Series sans tenir compte particulière de la disposition des données. Les seules exigences sont un nombre élevé d’IOPS, une faible latence et une fiabilité robuste. Bien que certaines pratiques d’excellence soient liées à des facteurs tels que la configuration RAID ou les pools de disques dynamiques, les projets EF-Series nécessitent rarement des modifications importantes de l’architecture de stockage globale pour exploiter ces fonctionnalités.

En revanche, la migration vers ONTAP nécessite généralement une plus grande considération de la disposition de la base de données pour s’assurer que la configuration finale offre une valeur maximale. À elle seule, ONTAP offre de nombreuses fonctionnalités pour un environnement de base de données, même sans effort d’architecture spécifique. Plus important encore, il permet de migrer vers un nouveau matériel sans interruption lorsque le matériel actuel arrive en fin de vie. De manière générale, une migration vers ONTAP est la dernière migration que vous auriez à effectuer. Ensuite, le matériel est mis à niveau et les données sont migrées sans interruption vers de nouveaux supports.

Avec une certaine planification, davantage d’avantages sont disponibles. Les considérations les plus importantes concernent l’utilisation des snapshots. Les copies Snapshot sont la base des sauvegardes, des restaurations et des opérations de clonage quasi-instantanées. Comme exemple de la puissance des snapshots, l’utilisation la plus répandue concerne une base de données unique de 996 To qui s’exécute sur environ 250 LUN sur 6 contrôleurs. Cette base de données peut être sauvegardée en 2 minutes, restaurée en 2 minutes et clonée en 15 minutes. Les autres avantages sont la capacité à déplacer les données au sein du cluster en réponse aux modifications des charges de travail et les contrôles de qualité de service (QoS) appliqués pour fournir de bonnes performances cohérentes dans un environnement à plusieurs bases de données.

Les technologies comme les contrôles de qualité de service, la relocalisation des données, la copie Snapshot et le clonage fonctionnent dans presque toutes les configurations. Cependant, certains pensent généralement être nécessaires pour maximiser les avantages. Dans certains cas, les dispositions du stockage de la base de données peuvent nécessiter des modifications de conception afin d’optimiser l’investissement dans la nouvelle baie de stockage. De telles modifications de conception peuvent avoir un impact sur la stratégie de migration, car les migrations basées sur les hôtes ou sur le stockage répliquent la disposition des données d’origine. Des étapes supplémentaires peuvent être nécessaires pour mener à bien la migration et assurer une disposition des données optimisée pour ONTAP. Les procédures indiquées à la "Présentation des procédures de migration Oracle" vous pouvez par la suite présenter certaines méthodes qui vous permettent non seulement de migrer une base de données, mais aussi de la migrer vers la configuration finale optimale en un minimum d’efforts.

Délai de mise en service

Vous devez déterminer la durée maximale autorisée de l'interruption de service pendant la mise en service. C'est une erreur courante de supposer que l'ensemble du processus de migration provoque des perturbations. De nombreuses tâches peuvent être effectuées avant le début d'une interruption de service, et de nombreuses options permettent d'effectuer la migration sans interruption ni panne. Même si une interruption est inévitable, vous devez toujours définir le temps d'interruption de service maximal autorisé, car la durée de la mise en service varie d'une procédure à l'autre.

Par exemple, la copie d'une base de données de 10 To prend généralement environ sept heures. Si l'entreprise a besoin d'une interruption de service de sept heures, la copie de fichiers est une option simple et sûre pour la migration. Si cinq heures sont inacceptables, un simple processus d'envoi de journaux (voir "Envoi de journaux Oracle") peut être configuré en déployant un minimum d'efforts afin de réduire le délai de mise en service à environ 15 minutes. Pendant ce temps, un administrateur de base de données peut terminer le processus. Si 15 ce n'est pas le cas, le processus de mise en service final peut être automatisé par script afin de réduire le délai de mise en service à quelques minutes seulement. Vous pouvez toujours accélérer une migration, mais cette opération a un coût en temps et en efforts. Les délais de mise en service doivent être déterminés en fonction des objectifs acceptables pour l'entreprise.

Chemin de retour arrière

Aucune migration n'est totalement sans risque. Même si la technologie fonctionne parfaitement, il y a toujours une possibilité d'erreur de l'utilisateur. Le risque associé au chemin de migration choisi doit être pris en compte parallèlement aux conséquences d'un échec de la migration. Par exemple, la fonctionnalité de migration transparente du stockage en ligne d'Oracle ASM est l'une de ses principales fonctionnalités, et cette méthode est l'une des plus fiables connues. Cependant, les données sont copiées de manière irréversible avec cette méthode. Dans le cas peu probable où un problème se produit avec ASM, il n'y a pas de chemin de sortie simple. La seule option consiste à restaurer l'environnement d'origine ou à utiliser ASM pour restaurer la migration vers les LUN d'origine. Le risque peut être réduit, mais pas éliminé, en effectuant une sauvegarde de type Snapshot sur le système de stockage d'origine, à condition que le système soit capable d'effectuer une telle opération.

Répétition

Certaines procédures de migration doivent être entièrement vérifiées avant leur exécution. La nécessité d'une migration et d'une répétition du processus de mise en service est courante dans les bases de données stratégiques pour lesquelles la migration doit réussir et où les temps d'indisponibilité doivent être minimisés. En outre, les tests d'acceptation par l'utilisateur sont fréquemment inclus dans le travail de post-migration et le système global ne peut être remis en production qu'une fois ces tests terminés.

S'il est nécessaire de répéter, plusieurs fonctionnalités ONTAP peuvent faciliter le processus. En particulier, les snapshots peuvent réinitialiser un environnement de test et créer rapidement plusieurs copies compactes d'un environnement de base de données.

Procédures

Présentation

De nombreuses procédures sont disponibles pour la migration d'une base de données Oracle. Le bon dépend des besoins de votre entreprise.

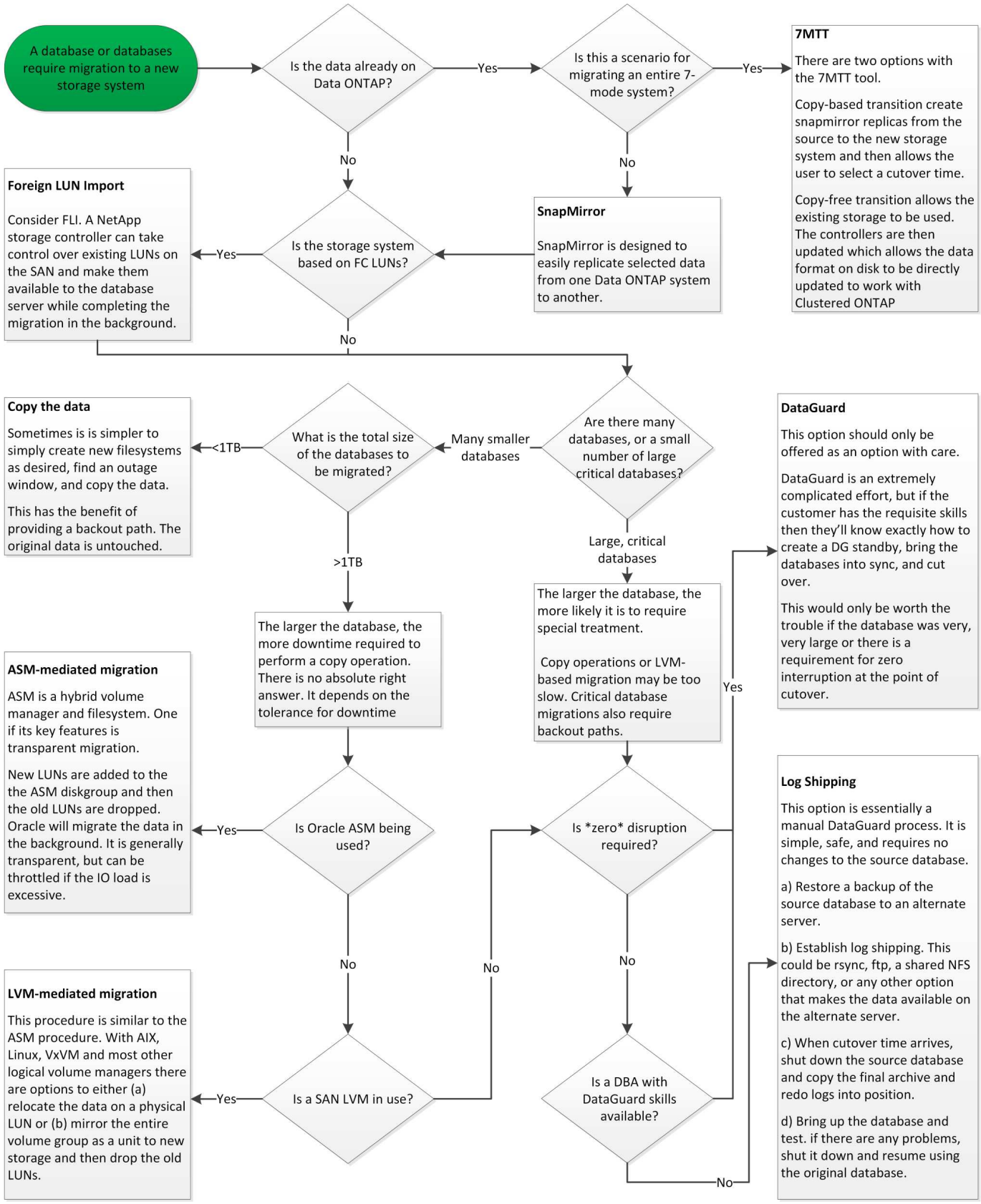
Dans de nombreux cas, les administrateurs système et les administrateurs de bases de données utilisent leurs propres méthodes de déplacement des données de volume physique, de mise en miroir et de déréplication, ou d'utilisation d'Oracle RMAN pour la copie des données.

Ces procédures sont fournies principalement à titre de conseils pour le personnel INFORMATIQUE qui connaît moins bien certaines des options disponibles. En outre, ces procédures illustrent les tâches, les exigences en termes de temps et les besoins en compétences de chaque approche de migration. Ainsi, d'autres parties, telles que NetApp et les services professionnels partenaires ou la direction INFORMATIQUE, peuvent mieux apprécier les exigences de chaque procédure.

Il n'existe pas de meilleure pratique unique pour créer une stratégie de migration. Pour créer un plan, il faut d'abord comprendre les options de disponibilité, puis sélectionner la méthode la mieux adaptée aux besoins de l'entreprise. La figure ci-dessous illustre les considérations de base et les conclusions types des clients, mais elle n'est pas universellement applicable à toutes les situations.

Par exemple, une étape soulève le problème de la taille totale de la base de données. L'étape suivante dépend si la base de données est supérieure ou inférieure à 1 To. Les étapes recommandées sont précisément des recommandations basées sur les pratiques standard des clients. La plupart des clients n'utiliseraient pas DataGuard pour copier une petite base de données, mais d'autres pourraient le faire. La plupart des clients ne tenteraient pas de copier une base de données de 50 To en raison du temps nécessaire, mais certaines peuvent disposer d'une fenêtre de maintenance suffisamment longue pour permettre une telle opération.

L'organigramme ci-dessous présente les différents éléments à prendre en compte pour déterminer le chemin de migration le plus approprié. Vous pouvez cliquer avec le bouton droit de la souris sur l'image et l'ouvrir dans un nouvel onglet pour améliorer la lisibilité.



.

Déplacement du fichier de données en ligne

Oracle 12cR1 et versions supérieures incluent la possibilité de déplacer un fichier de données pendant que la base de données reste en ligne. Il fonctionne en outre entre différents types de systèmes de fichiers. Par

exemple, un fichier de données peut être déplacé d'un système de fichiers xfs vers ASM. Cette méthode n'est généralement pas utilisée à grande échelle en raison du nombre d'opérations de déplacement de fichiers de données individuelles qui seraient requises. Toutefois, il est important de tenir compte de cette méthode avec des bases de données plus petites et moins de fichiers de données.

En outre, le simple déplacement d'un fichier de données est une bonne option pour migrer des parties de bases de données existantes. Par exemple, les fichiers de données moins actifs peuvent être transférés vers un stockage plus économique, tel qu'un volume FabricPool qui peut stocker les blocs inactifs dans le magasin d'objets.

Migration au niveau de la base de données

La migration au niveau de la base de données signifie que la base de données peut déplacer des données. Plus précisément, cela signifie l'envoi de journaux. Des technologies telles que RMAN et ASM sont des produits Oracle, mais pour la migration, elles fonctionnent au niveau de l'hôte où elles copient les fichiers et gèrent les volumes.

Envoi de journaux

La base de la migration au niveau de la base de données est le journal d'archivage Oracle, qui contient un journal des modifications apportées à la base de données. La plupart du temps, un journal d'archivage fait partie d'une stratégie de sauvegarde et de restauration. Le processus de restauration commence par la restauration d'une base de données, puis la relecture d'un ou plusieurs journaux d'archivage pour ramener la base de données à l'état souhaité. Cette même technologie de base peut être utilisée pour effectuer une migration avec une interruption des opérations nulle ou minime. Plus important encore, cette technologie permet la migration tout en conservant la base de données d'origine intacte, ce qui permet de conserver un chemin de retour.

Le processus de migration commence par la restauration d'une sauvegarde de base de données sur un serveur secondaire. Vous pouvez le faire de différentes manières, mais la plupart des clients utilisent leur application de sauvegarde normale pour restaurer les fichiers de données. Une fois les fichiers de données restaurés, les utilisateurs établissent une méthode d'envoi des journaux. L'objectif est de créer un flux constant de journaux d'archivage générés par la base de données primaire et de les relire sur la base de données restaurée afin de les conserver dans un état similaire. Lorsque le délai de mise en service arrive, la base de données source est complètement arrêtée et les journaux d'archivage finaux, et dans certains cas les journaux de reprise, sont copiés et relus. Il est essentiel que les journaux de reprise soient également pris en compte, car ils peuvent contenir certaines des transactions finales validées.

Une fois ces journaux transférés et relus, les deux bases de données sont cohérentes l'une avec l'autre. À ce stade, la plupart des clients effectuent des tests de base. Si des erreurs sont commises pendant le processus de migration, la relecture du journal doit signaler les erreurs et échouer. Il est toujours conseillé d'effectuer des tests rapides basés sur des requêtes connues ou des activités applicatives pour vérifier que la configuration est optimale. Il est également courant de créer une table de test finale avant d'arrêter la base de données d'origine pour vérifier qu'elle est présente dans la base de données migrée. Cette étape permet de s'assurer qu'aucune erreur n'a été effectuée lors de la synchronisation finale du journal.

Une simple migration d'envoi de journaux peut être configurée hors bande par rapport à la base de données d'origine, ce qui la rend particulièrement utile pour les bases de données stratégiques. Il n'est pas nécessaire de modifier la configuration de la base de données source, car la restauration et la configuration initiale de l'environnement de migration n'affectent pas les opérations de production. Une fois l'envoi de journaux configuré, il impose des demandes d'E/S sur les serveurs de production. Cependant, l'envoi de journaux se compose de simples lectures séquentielles des journaux d'archivage, qui n'ont probablement aucun impact sur les performances des bases de données de production.

L'expédition de journaux s'est avérée particulièrement utile pour les projets de migration longue distance à

taux de changement élevé. Dans un cas, une seule base de données de 220 To a été migrée vers un nouvel emplacement situé à environ 500 kilomètres. Le taux de modification était extrêmement élevé et les restrictions de sécurité empêchaient l'utilisation d'une connexion réseau. L'expédition des journaux a été effectuée à l'aide de bandes et de coursiers. Une copie de la base de données source a d'abord été restaurée à l'aide des procédures décrites ci-dessous. Les journaux ont ensuite été expédiés chaque semaine par messagerie jusqu'au moment de la mise en service, lorsque le jeu final de bandes a été livré et que les journaux ont été appliqués à la base de données de réplica.

Oracle DataGuard

Dans certains cas, un environnement DataGuard complet est garanti. Il est incorrect d'utiliser le terme DataGuard pour faire référence à toute configuration d'envoi de journaux ou de base de données de secours. Oracle DataGuard est un framework complet de gestion de la réplication de base de données, mais il ne s'agit pas d'une technologie de réplication. Le principal avantage d'un environnement DataGuard complet dans un effort de migration est le basculement transparent d'une base de données à une autre. DataGuard permet également un basculement transparent vers la base de données d'origine en cas de problème, tel qu'un problème de performances ou de connectivité réseau avec le nouvel environnement. Un environnement DataGuard entièrement configuré nécessite la configuration non seulement de la couche de base de données, mais aussi des applications pour que les applications puissent détecter un changement dans l'emplacement de la base de données primaire. En général, il n'est pas nécessaire d'utiliser DataGuard pour effectuer une migration, mais certains clients possèdent une expertise DataGuard étendue en interne et en dépendent déjà pour le travail de migration.

Architecture

Comme évoqué précédemment, l'exploitation des fonctionnalités avancées des baies de stockage nécessite parfois de modifier l'organisation de la base de données. De plus, une modification du protocole de stockage, telle que le passage d'ASM à un système de fichiers NFS, modifie nécessairement la disposition du système de fichiers.

L'un des principaux avantages des méthodes d'envoi de journaux, y compris DataGuard, est que la destination de réplication ne doit pas correspondre à la source. Il n'y a pas de problème avec l'utilisation d'une approche d'envoi de journaux pour migrer d'ASM vers un système de fichiers standard, et inversement. La disposition précise des fichiers de données peut être modifiée à la destination pour optimiser l'utilisation de la technologie de base de données enfichable (PDB) ou pour définir des contrôles QoS de manière sélective sur certains fichiers. En d'autres termes, un processus de migration basé sur l'envoi de journaux vous permet d'optimiser facilement et en toute sécurité l'organisation du stockage de la base de données.

Ressources du serveur

La migration au niveau de la base de données est limitée par le besoin d'un second serveur. Ce second serveur peut être utilisé de deux manières :

1. Vous pouvez utiliser le second serveur comme nouveau domicile permanent pour la base de données.
2. Vous pouvez utiliser le second serveur comme serveur temporaire de transfert. Une fois la migration des données vers la nouvelle baie de stockage terminée et testée, les systèmes de fichiers LUN ou NFS sont déconnectés du serveur intermédiaire et reconnectés au serveur d'origine.

La première option est la plus simple, mais son utilisation peut ne pas être possible dans les environnements très vastes nécessitant des serveurs très puissants. La deuxième option nécessite un travail supplémentaire pour replacer les systèmes de fichiers à leur emplacement d'origine. Il peut s'agir d'une opération simple dans laquelle NFS est utilisé comme protocole de stockage car les systèmes de fichiers peuvent être démontés du serveur de transfert et remontés sur le serveur d'origine.

Les systèmes de fichiers basés sur les blocs nécessitent un travail supplémentaire pour mettre à jour le zoning FC ou les initiateurs iSCSI. Avec la plupart des gestionnaires de volumes logiques (y compris ASM), les LUN sont automatiquement détectées et mises en ligne après leur mise à disposition sur le serveur d'origine. Cependant, certaines implémentations de système de fichiers et de LVM peuvent nécessiter davantage de travail pour exporter et importer les données. La procédure précise peut varier, mais il est généralement facile d'établir une procédure simple et reproductible pour terminer la migration et réexécuter les données sur le serveur d'origine.

Bien qu'il soit possible de configurer l'envoi de journaux et de répliquer une base de données dans un environnement de serveur unique, la nouvelle instance doit avoir un SID de processus différent pour pouvoir relire les journaux. Il est possible d'afficher temporairement la base de données sous un autre ensemble d'ID de processus avec un SID différent et de la modifier ultérieurement. Toutefois, cela peut entraîner de nombreuses activités de gestion complexes et mettre l'environnement de base de données en danger d'erreur de la part des utilisateurs.

Migration au niveau de l'hôte

La migration des données au niveau de l'hôte implique l'utilisation du système d'exploitation hôte et des utilitaires associés pour terminer la migration. Ce processus inclut tout utilitaire qui copie les données, y compris Oracle RMAN et Oracle ASM.

Copie de données

La valeur d'une opération de copie simple ne doit pas être sous-estimée. Les infrastructures réseau modernes peuvent déplacer des données à un taux de gigaoctets par seconde. Les opérations de copie de fichiers reposent sur des E/S efficaces en lecture et écriture séquentielles Si une opération de copie de l'hôte est plus perturbant que l'envoi de journaux, la migration ne se limite pas au déplacement des données. Elle inclut généralement les modifications apportées au réseau, au délai de redémarrage de la base de données et aux tests de post-migration.

Le temps réel nécessaire à la copie des données peut ne pas être important. En outre, une opération de copie préserve un chemin de retour garanti, car les données d'origine ne sont pas modifiées. En cas de problème pendant le processus de migration, les systèmes de fichiers d'origine avec les données d'origine peuvent être réactivés.

Changement de plate-forme

Le changement de plate-forme fait référence à un changement de type de CPU. Lorsqu'une base de données est migrée d'une plate-forme Solaris, AIX ou HP-UX traditionnelle vers Linux x86, les données doivent être reformatées en raison de modifications de l'architecture CPU. Les processeurs SPARC, IA64 et POWER sont connus sous le nom de processeurs big endian, tandis que les architectures x86 et x86_64 sont connues sous le nom de Little endian. Par conséquent, certaines données des fichiers de données Oracle sont triées différemment selon le processeur utilisé.

Jusqu'ici, les clients ont généralement utilisé DataPump pour répliquer des données sur plusieurs plateformes. DataPump est un utilitaire qui crée un type spécial d'exportation de données logiques qui peut être importé plus rapidement dans la base de données de destination. Comme il crée une copie logique des données, DataPump laisse derrière lui les dépendances de l'endianness du processeur. DataPump est encore utilisé par certains clients pour le changement de plateforme, mais une option plus rapide est désormais disponible avec Oracle 11g : les tablespaces interplateformes transportables. Cette avance permet de convertir un espace de table en un format endian différent. Il s'agit d'une transformation physique qui offre de meilleures performances qu'une exportation DataPump, qui doit convertir les octets physiques en données logiques, puis les convertir en octets physiques.

Une discussion complète sur DataPump et les tablespaces transportables va au-delà de la documentation

NetApp portée, mais NetApp propose quelques recommandations basées sur notre expérience d'assistance aux clients lors de la migration vers une nouvelle baie de stockage dans le cadre d'une nouvelle architecture de processeur :

- Si DataPump est utilisé, le temps nécessaire à la migration doit être mesuré dans un environnement de test. Les clients sont parfois surpris du temps nécessaire à la réalisation de la migration. Cette interruption supplémentaire imprévue peut provoquer des interruptions.
- De nombreux clients pensent à tort que les tablespaces transportables multi plates-formes ne nécessitent pas de conversion de données. Lorsqu'une CPU avec un autre endian est utilisée, un RMAN `convert` l'opération doit être effectuée au préalable sur les fichiers de données. Cette opération n'est pas instantanée. Dans certains cas, le processus de conversion peut être accéléré en ayant plusieurs threads fonctionnant sur différents fichiers de données, mais le processus de conversion ne peut pas être évité.

Migration basée sur le gestionnaire de volumes logiques

Les LVM fonctionnent en déregroupant un groupe d'une ou de plusieurs LUN en petites unités généralement appelées extensions. Le pool d'extensions est ensuite utilisé comme source pour créer des volumes logiques qui sont essentiellement virtualisés. Cette couche de virtualisation apporte de la valeur de plusieurs manières :

- Les volumes logiques peuvent utiliser des extensions tirées de plusieurs LUN. Lorsqu'un système de fichiers est créé sur un volume logique, il peut exploiter les performances maximales de toutes les LUN. Il favorise également le chargement homogène de toutes les LUN du groupe de volumes, pour des performances plus prévisibles.
- Les volumes logiques peuvent être redimensionnés en ajoutant et, dans certains cas, en supprimant des extensions. Le redimensionnement d'un système de fichiers sur un volume logique s'effectue généralement sans interruption.
- Le déplacement des extensions sous-jacentes permet de migrer les volumes logiques sans interruption.

La migration à l'aide d'un LVM fonctionne de deux manières : déplacer une extension ou mettre en miroir/démirroring une extension. La migration des LVM utilise des E/S séquentielles de blocs de grande taille efficaces et pose rarement des problèmes de performances. Si ce problème survient, il existe généralement des options pour limiter le taux d'E/S. Cela augmente le temps nécessaire à la migration, tout en réduisant la charge d'E/S sur l'hôte et les systèmes de stockage.

Miroir et démiroir

Certains gestionnaires de volumes, tels que AIX LVM, permettent à l'utilisateur de spécifier le nombre de copies pour chaque extension et de contrôler les périphériques qui hébergent chaque copie. La migration s'effectue par la mise en miroir d'un volume logique existant sur les extensions sous-jacentes des nouveaux volumes, l'attente de la synchronisation des copies, puis l'abandon de l'ancienne copie. Si un chemin de retour arrière est souhaité, un instantané des données d'origine peut être créé avant le point de suppression de la copie miroir. Il est également possible d'arrêter brièvement le serveur pour masquer les LUN d'origine avant de forcer la suppression des copies miroir contenues. Cela permet de conserver une copie récupérable des données à leur emplacement d'origine.

Migration d'extension

La plupart des gestionnaires de volumes permettent la migration des extensions, et il arrive parfois que plusieurs options existent. Par exemple, certains gestionnaires de volumes permettent à un administrateur de déplacer les extensions individuelles d'un volume logique spécifique de l'ancien vers le nouveau stockage. Les gestionnaires de volumes tels que Linux LVM2 offrent le `pvmove` Qui déplace toutes les extensions du périphérique LUN spécifié vers une nouvelle LUN. Une fois l'ancien LUN évacué, il est possible de le retirer.

> Le risque principal pour les opérations est la suppression des anciennes LUN inutilisées de la configuration. Une attention toute particulière doit être portée au changement de segmentation FC et au retrait des périphériques LUN obsolètes.

Gestion automatique du stockage par Oracle

Oracle ASM est un gestionnaire de volumes logiques et un système de fichiers combinés. À un niveau élevé, Oracle ASM prend un ensemble de LUN, les répartit en petites unités d'allocation et les présente comme un seul volume appelé groupe de disques ASM. ASM permet également de mettre en miroir le groupe de disques en définissant le niveau de redondance. Un volume peut être sans miroir (redondance externe), en miroir (redondance normale) ou en miroir tridirectionnel (redondance élevée). La configuration du niveau de redondance doit être effectuée avec précaution car il ne peut pas être modifié après sa création.

ASM fournit également des fonctionnalités de système de fichiers. Bien que le système de fichiers ne soit pas visible directement depuis l'hôte, la base de données Oracle peut créer, déplacer et supprimer des fichiers et des répertoires sur un groupe de disques ASM. Vous pouvez également naviguer dans la structure à l'aide de l'utilitaire asmcmd.

Comme pour les autres implémentations LVM, Oracle ASM optimise les performances d'E/S en segmentant et en équilibrant les E/S de chaque fichier sur l'ensemble des LUN disponibles. Deuxièmement, les extensions sous-jacentes peuvent être déplacées pour permettre le redimensionnement du groupe de disques ASM ainsi que la migration. Oracle ASM automatise le processus tout au long de l'opération de rééquilibrage. Les nouvelles LUN sont ajoutées à un groupe de disques ASM et les anciennes LUN sont abandonnées, ce qui déclenche le déplacement d'extension et le DROP suivant de la LUN évacuée du groupe de disques. Ce processus est l'une des méthodes de migration les plus éprouvées, et la fiabilité d'ASM pour assurer une migration transparente est probablement sa fonctionnalité la plus importante.

> Comme le niveau de mise en miroir d'Oracle ASM est fixe, il ne peut pas être utilisé avec la méthode de migration miroir et démiroir.

Migration au niveau du stockage

La migration au niveau du stockage implique d'effectuer la migration au-dessous des niveaux des applications et du système d'exploitation. Auparavant, il fallait parfois utiliser des périphériques spécialisés qui copiaient les LUN au niveau du réseau, mais ces fonctionnalités sont désormais natives dans ONTAP.

SnapMirror

La migration de bases de données entre des systèmes NetApp est presque effectuée de manière universelle avec le logiciel de réplication des données NetApp SnapMirror. Ce processus implique la configuration d'une relation de miroir pour les volumes à migrer, leur permettant ainsi de se synchroniser, puis d'attendre la fenêtre de mise en service. Lorsqu'elle arrive, la base de données source est arrêtée, une dernière mise à jour miroir est effectuée et le miroir est cassé. Les volumes de réplica sont alors prêts à l'emploi, soit en montant un répertoire de système de fichiers NFS contenu, soit en découvrant les LUN contenues et en démarrant la base de données.

La relocalisation des volumes dans un seul cluster ONTAP n'est pas considérée comme une migration, mais plutôt comme une routine `volume move` fonctionnement. SnapMirror est utilisé en tant que moteur de réplication des données au sein du cluster. Ce processus est entièrement automatisé. Il n'y a pas d'étape de migration supplémentaire à effectuer lorsque les attributs du volume, tels que le mappage de LUN ou les autorisations d'exportation NFS, sont déplacés avec le volume lui-même. La relocalisation ne prend pas en charge l'hôte. Dans certains cas, il convient de mettre à jour l'accès au réseau pour s'assurer que les données nouvellement déplacées sont accessibles de la manière la plus efficace possible, mais sans interruption.

Importation de LUN étrangères (FLI)

La FLI est une fonctionnalité qui permet à un système Data ONTAP exécutant la version 8.3 ou supérieure de migrer un LUN existant à partir d'une autre baie de stockage. La procédure est simple : le système ONTAP est zoné sur la baie de stockage existante comme s'il s'agissait d'un autre hôte SAN. Data ONTAP prend alors le contrôle des LUN héritées souhaitées et migre les données sous-jacentes. De plus, le processus d'importation utilise les paramètres d'efficacité du nouveau volume lors de la migration des données. Ainsi, les données peuvent être compressées et dédupliquées en ligne pendant le processus de migration.

La première implémentation de FLI dans Data ONTAP 8.3 a permis uniquement la migration hors ligne. Ce transfert était extrêmement rapide, mais cela signifiait que les données de LUN étaient indisponibles jusqu'à la fin de la migration. La migration en ligne a été introduite dans Data ONTAP 8.3.1. Ce type de migration minimise les interruptions en permettant à ONTAP de transmettre des données LUN lors du processus de transfert. Il y a une brève interruption lors de la remise en place de l'hôte pour l'utilisation des LUN via ONTAP. Cependant, dès que ces modifications sont apportées, les données sont de nouveau accessibles et restent accessibles tout au long du processus de migration.

Les E/S de lecture sont proxées via ONTAP jusqu'à la fin de l'opération de copie, tandis que les E/S d'écriture sont écrites de manière synchrone sur les LUN étrangères et ONTAP. Les deux copies LUN sont ainsi synchronisées jusqu'à ce que l'administrateur exécute une mise en service complète qui libère le LUN étranger et ne réplique plus les écritures.

FLI est conçu pour fonctionner avec FC. Toutefois, si vous souhaitez passer à iSCSI, le LUN migré peut facilement être remappé en tant que LUN iSCSI une fois la migration terminée.

Parmi les caractéristiques de FLI figurent la détection et le réglage automatiques de l'alignement. Dans ce contexte, le terme alignement fait référence à une partition sur un périphérique LUN. Pour des performances optimales, les E/S doivent être alignées sur des blocs de 4 Ko. Si une partition est placée à un décalage qui n'est pas un multiple de 4K, les performances en pâtissent.

Il existe un deuxième aspect de l'alignement qui ne peut pas être corrigé en réglant un décalage de partition, c'est-à-dire la taille du bloc du système de fichiers. Par exemple, un système de fichiers ZFS prend généralement par défaut une taille de bloc interne de 512 octets. D'autres clients utilisant AIX ont parfois créé des systèmes de fichiers jfs2 avec une taille de bloc de 512 ou 1, 024 octets. Bien que le système de fichiers puisse être aligné sur une limite de 4 Ko, les fichiers créés dans ce système de fichiers ne le sont pas et les performances en pâtissent.

FLI ne doit pas être utilisé dans ces circonstances. Bien que les données soient accessibles après la migration, vous obtenez des systèmes de fichiers avec de graves limitations de performances. En principe, tout système de fichiers prenant en charge une charge de travail de remplacement aléatoire sur ONTAP doit utiliser une taille de bloc de 4 Ko. Cela s'applique principalement aux charges de travail telles que les fichiers de données de base de données et les déploiements VDI. La taille de bloc peut être identifiée à l'aide des commandes appropriées du système d'exploitation hôte.

Par exemple, sous AIX, la taille de bloc peut être affichée avec `lsfs -q`. Avec Linux, `xfs_info` et `tune2fs` peut être utilisé pour `xfs` et `ext3/ext4`, respectivement. Avec `zfs`, la commande est `zdb -C`.

Le paramètre qui contrôle la taille du bloc est `ashift` et la valeur par défaut est généralement 9, soit 2^9, ou 512 octets. Pour des performances optimales, le `ashift` La valeur doit être 12 (2^12=4K). Cette valeur est définie au moment de la création du zpool et ne peut pas être modifiée, ce qui signifie que les zpools de données avec un `ashift` une migration autre que 12 doit être effectuée en copiant les données vers un nouveau zpool.

Oracle ASM n'a pas de taille de bloc fondamentale. La seule exigence est que la partition sur laquelle le disque ASM est construit doit être correctement alignée.

Outil de transition 7-mode

L'outil 7-mode transition Tool (7MTT) est un utilitaire d'automatisation utilisé pour migrer de grandes configurations 7-mode vers ONTAP. La plupart des clients de bases de données trouvent d'autres méthodes plus faciles, notamment parce qu'ils migrent généralement leurs environnements de bases de données par base de données plutôt que de déplacer l'intégralité de l'empreinte du stockage. De plus, les bases de données ne font souvent partie que d'un environnement de stockage plus important. Les bases de données sont donc souvent migrées individuellement, puis le reste de l'environnement peut être déplacé avec 7MTT.

Les clients sont de petite taille, mais nombreux. Ils disposent de systèmes de stockage dédiés à des environnements de base de données complexes. Ces environnements peuvent contenir de nombreux volumes, snapshots et de nombreuses informations de configuration telles que les autorisations d'exportation, les groupes initiateurs de LUN, les autorisations utilisateur et la configuration du protocole d'accès aux répertoires légers. Dans de tels cas, les fonctionnalités d'automatisation de l'outil 7MTT simplifient considérablement la migration.

7MTT peut fonctionner dans deux modes :

- **Transition basée sur les copies (CBT).** dans le nouvel environnement, l'outil 7MTT avec CBT configure les volumes SnapMirror à partir d'un système 7- mode existant. Une fois les données synchronisées, l'outil 7MTT orchestre le processus de mise en service.
- **Transition sans copie.** 7MTT avec la transition sans copie repose sur la conversion des tiroirs disques 7-mode existants sans déplacement des données. Aucune donnée n'est copiée et les tiroirs disques existants peuvent être réutilisés. La protection des données et la configuration de l'efficacité du stockage existantes sont préservées.

La différence principale entre ces deux options est que la transition sans copie constitue une approche globale où tous les tiroirs disques rattachés à la paire HA 7-mode d'origine doivent être transférés vers le nouvel environnement. Il n'existe aucune option pour déplacer un sous-ensemble de tiroirs. L'approche basée sur les copies permet de déplacer des volumes sélectionnés. Par ailleurs, une fenêtre de mise en service peut être plus longue et la transition sans copie est liée à l'alignement des tiroirs disques et à la conversion des métadonnées. En fonction de son expérience sur le terrain, NetApp recommande de consacrer 1 heure au déplacement et à la réinstallation des tiroirs disques, et entre 15 minutes et 2 heures à la conversion des métadonnées.

Migration des fichiers de données

Vous pouvez déplacer individuellement les fichiers de données Oracle via une seule commande.

Par exemple, la commande suivante déplace le fichier de données IOPST.dbf du système de fichiers `/oradata2` vers le système de fichiers `/oradata3`.

```
SQL> alter database move datafile  '/oradata2/NTAP/IOPS002.dbf' to
'/oradata3/NTAP/IOPS002.dbf';
Database altered.
```

Le déplacement d'un fichier de données avec cette méthode peut être lent, mais il ne doit normalement pas produire suffisamment d'E/S pour interférer avec les charges de travail quotidiennes des bases de données. En revanche, la migration via le rééquilibrage d'ASM peut s'exécuter beaucoup plus rapidement, mais au détriment du ralentissement de la base de données globale pendant le déplacement des données.

Le temps nécessaire à la migration des fichiers de données peut être mesuré en créant un fichier de données de test et en le déplaçant. Le temps écoulé pour l'opération est enregistré dans les données v$session :

```
SQL> set linesize 300;
SQL> select elapsed_seconds||':'||message from v$session_longops;
ELAPSED_SECONDS||':'||MESSAGE
------------------------------------------------------------------------
---------------
351:Online data file move: data file 8: 22548578304 out of 22548578304
bytes done
SQL> select bytes / 1024 / 1024 /1024 as GB from dba_data_files where
FILE_ID = 8;
        GB
----------
        21
```

Dans cet exemple, le fichier déplacé était le fichier de données 8, dont la taille était de 21 Go et dont la migration nécessitait environ 6 minutes. Le temps nécessaire dépend évidemment des capacités du système de stockage, du réseau de stockage et de l'activité globale de la base de données au moment de la migration.

Envoi de journaux

L'objectif d'une migration à l'aide de l'envoi de journaux est de créer une copie des fichiers de données d'origine à un nouvel emplacement, puis d'établir une méthode d'expédition des modifications dans le nouvel environnement.

Une fois établie, l'envoi et la relecture des journaux peuvent être automatisés afin de maintenir la base de données de réplica largement synchronisée avec la source. Par exemple, une tâche cron peut être planifiée pour (a) copier les journaux les plus récents vers le nouvel emplacement et (b) les relire toutes les 15 minutes. L'interruption au moment de la mise en service est ainsi minimale, car la lecture des journaux d'archivage ne doit pas dépasser 15 minutes.

La procédure présentée ci-dessous est également essentiellement une opération de clonage de base de données. La logique illustrée est similaire au moteur de NetApp SnapManager pour Oracle (SMO) et du plug-in Oracle NetApp SnapCenter. Certains clients ont utilisé la procédure présentée dans des scripts ou des workflows WFA pour des opérations de clonage personnalisé. Bien que cette procédure soit plus manuelle qu'avec SMO ou SnapCenter, elle reste facilement scriptée, et les API de gestion des données de ONTAP simplifient davantage le processus.

Envoi de journaux - système de fichiers vers le système de fichiers

Cet exemple illustre la migration d'une base de données appelée WAFFLE d'un système de fichiers ordinaire vers un autre système de fichiers ordinaire situé sur un serveur différent. Il illustre également l'utilisation de SnapMirror pour effectuer une copie rapide des fichiers de données, mais cela ne fait pas partie intégrante de la procédure globale.

Créer une sauvegarde de base de données

La première étape consiste à créer une sauvegarde de base de données. Plus précisément, cette procédure nécessite un ensemble de fichiers de données pouvant être utilisés pour la relecture des journaux d'archivage.

De production

Dans cet exemple, la base de données source se trouve sur un système ONTAP. La méthode la plus simple pour créer une sauvegarde d'une base de données consiste à utiliser un instantané. La base de données est placée en mode de sauvegarde à chaud pendant quelques secondes `snapshot create` l'opération est exécutée sur le volume hébergeant les fichiers de données.

```
SQL> alter database begin backup;
Database altered.
```

```
Cluster01::*> snapshot create -vserver vserver1 -volume jfsc1_oradata
hotbackup
Cluster01::*>
```

```
SQL> alter database end backup;
Database altered.
```

Le résultat est un instantané sur le disque appelé `hotbackup` qui contient une image des fichiers de données en mode de sauvegarde à chaud. Lorsqu'elles sont combinées avec les journaux d'archivage appropriés pour assurer la cohérence des fichiers de données, les données de cet instantané peuvent servir de base à une restauration ou à un clone. Dans ce cas, il est répliqué sur le nouveau serveur.

Restaurer dans un nouvel environnement

La sauvegarde doit maintenant être restaurée dans le nouvel environnement. Cette opération peut être effectuée de plusieurs façons, notamment Oracle RMAN, la restauration à partir d'une application de sauvegarde comme NetBackup ou une simple opération de copie des fichiers de données placés en mode de sauvegarde à chaud.

Dans cet exemple, SnapMirror est utilisé pour répliquer la sauvegarde à chaud de snapshot vers un nouvel emplacement.

1. Créez un volume pour recevoir les données de snapshot. Initialiser la mise en miroir à partir de `jfsc1_oradata` à `vol_oradata`.

   ```
   Cluster01::*> volume create -vserver vserver1 -volume vol_oradata
   -aggregate data_01 -size 20g -state online -type DP -snapshot-policy
   none -policy jfsc3
   [Job 833] Job succeeded: Successful
   ```

```
Cluster01::*> snapmirror initialize -source-path vserver1:jfsc1_oradata
-destination-path vserver1:vol_oradata
Operation is queued: snapmirror initialize of destination
"vserver1:vol_oradata".
Cluster01::*> volume mount -vserver vserver1 -volume vol_oradata
-junction-path /vol_oradata
Cluster01::*>
```

2. Une fois l'état défini par SnapMirror, indiquant que la synchronisation est terminée, mettre à jour le miroir en fonction du snapshot souhaité.

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields state
source-path             destination-path        state
----------------------- ----------------------- ------------
vserver1:jfsc1_oradata vserver1:vol_oradata SnapMirrored
```

```
Cluster01::*> snapmirror update -destination-path vserver1:vol_oradata
-source-snapshot hotbackup
Operation is queued: snapmirror update of destination
"vserver1:vol_oradata".
```

3. La synchronisation peut être vérifiée en affichant le `newest-snapshot` champ sur le volume miroir.

```
Cluster01::*> snapmirror show -destination-path vserver1:vol_oradata
-fields newest-snapshot
source-path             destination-path        newest-snapshot
----------------------- ----------------------- ---------------
vserver1:jfsc1_oradata vserver1:vol_oradata hotbackup
```

4. Le miroir peut alors être cassé.

```
Cluster01::> snapmirror break -destination-path vserver1:vol_oradata
Operation succeeded: snapmirror break for destination
"vserver1:vol_oradata".
Cluster01::>
```

5. Montez le nouveau système de fichiers.avec les systèmes de fichiers en mode bloc, les procédures précises varient en fonction du LVM utilisé. Le zoning FC ou les connexions iSCSI doivent être configurés. Une fois la connectivité aux LUN établie, des commandes telles que Linux `pvscan` Il peut être nécessaire de déterminer quels groupes de volumes ou LUN doivent être configurés correctement pour être détectables par ASM.

Dans cet exemple, un simple système de fichiers NFS est utilisé. Ce système de fichiers peut être monté directement.

```
fas8060-nfs1:/vol_oradata        19922944   1639360   18283584   9%
/oradata
fas8060-nfs1:/vol_logs            9961472       128    9961344   1%
/logs
```

Créer un modèle de création de fichier de contrôle

Vous devez ensuite créer un modèle de fichier de contrôle. Le `backup controlfile to trace` commande crée des commandes texte pour recréer un fichier de contrôle. Dans certaines circonstances, cette fonction peut être utile pour restaurer une base de données à partir d'une sauvegarde, et elle est souvent utilisée avec des scripts qui effectuent des tâches telles que le clonage de base de données.

1. Le résultat de la commande suivante est utilisé pour recréer les fichiers de contrôle pour la base de données migrée.

```
SQL> alter database backup controlfile to trace as '/tmp/waffle.ctrl';
Database altered.
```

2. Une fois les fichiers de contrôle créés, copiez-les sur le nouveau serveur.

```
[oracle@jfsc3 tmp]$ scp oracle@jfsc1:/tmp/waffle.ctrl /tmp/
oracle@jfsc1's password:
waffle.ctrl                                           100% 5199
5.1KB/s   00:00
```

Sauvegarde du fichier de paramètres

Un fichier de paramètres est également requis dans le nouvel environnement. La méthode la plus simple consiste à créer un fichier pfile à partir du fichier spfile ou pfile actuel. Dans cet exemple, la base de données source utilise un fichier spfile.

```
SQL> create pfile='/tmp/waffle.tmp.pfile' from spfile;
File created.
```

Créer une entrée oratab

La création d'une entrée oratab est requise pour le bon fonctionnement des utilitaires tels que oraenv. Pour créer une entrée oratab, procédez comme suit.

```
WAFFLE:/orabin/product/12.1.0/dbhome_1:N
```

Préparer la structure du répertoire

Si les répertoires requis n'étaient pas déjà présents, vous devez les créer ou la procédure de démarrage de la base de données échoue. Pour préparer la structure de répertoires, remplissez les conditions minimales suivantes.

```
[oracle@jfsc3 ~]$ . oraenv
ORACLE_SID = [oracle] ? WAFFLE
The Oracle base has been set to /orabin
[oracle@jfsc3 ~]$ cd $ORACLE_BASE
[oracle@jfsc3 orabin]$ cd admin
[oracle@jfsc3 admin]$ mkdir WAFFLE
[oracle@jfsc3 admin]$ cd WAFFLE
[oracle@jfsc3 WAFFLE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

Mises à jour du fichier de paramètres

1. Pour copier le fichier de paramètres sur le nouveau serveur, exécutez les commandes suivantes. L'emplacement par défaut est le `$ORACLE_HOME/dbs` répertoire. Dans ce cas, le fichier pfile peut être placé n'importe où. Il est utilisé uniquement comme étape intermédiaire dans le processus de migration.

```
[oracle@jfsc3 admin]$ scp oracle@jfsc1:/tmp/waffle.tmp.pfile
$ORACLE_HOME/dbs/waffle.tmp.pfile
oracle@jfsc1's password:
waffle.pfile                                          100%  916
0.9KB/s   00:00
```

1. Modifiez le fichier selon vos besoins. Par exemple, si l'emplacement du journal d'archive a changé, le fichier pfile doit être modifié pour refléter le nouvel emplacement. Dans cet exemple, seuls les fichiers de contrôle sont déplacés, en partie pour les distribuer entre les systèmes de fichiers journaux et de données.

```
[root@jfsc1 tmp]# cat waffle.pfile
WAFFLE.__data_transfer_cache_size=0
WAFFLE.__db_cache_size=507510784
WAFFLE.__java_pool_size=4194304
WAFFLE.__large_pool_size=20971520
WAFFLE.__oracle_base='/orabin'#ORACLE_BASE set from environment
WAFFLE.__pga_aggregate_target=268435456
WAFFLE.__sga_target=805306368
WAFFLE.__shared_io_pool_size=29360128
WAFFLE.__shared_pool_size=234881024
WAFFLE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/WAFFLE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/oradata//WAFFLE/control01.ctl','/oradata//WAFFLE/contr
ol02.ctl'
*.control_files='/oradata/WAFFLE/control01.ctl','/logs/WAFFLE/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='WAFFLE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=WAFFLEXDB)'
*.log_archive_dest_1='LOCATION=/logs/WAFFLE/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

2. Une fois les modifications terminées, créez un fichier spfile basé sur ce fichier pfile.

```
SQL> create spfile from pfile='waffle.tmp.pfile';
File created.
```

Recréer les fichiers de contrôle

Dans une étape précédente, la sortie de `backup controlfile to trace` a été copié sur le nouveau serveur. La partie spécifique de la sortie requise est le `controlfile recreation` commande. Ces informations se trouvent dans le fichier sous la section marquée `Set #1. NORESETLOGS`. Il commence par la ligne `create controlfile reuse database` et doit inclure le mot `noresetlogs`. Il se termine par le caractère point-virgule (;).

1. Dans cet exemple de procédure, le fichier se lit comme suit.

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/logs/WAFFLE/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/logs/WAFFLE/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/logs/WAFFLE/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

2. Modifiez ce script comme vous le souhaitez pour refléter le nouvel emplacement des différents fichiers. Par exemple, certains fichiers de données connus pour prendre en charge des E/S élevées peuvent être redirigés vers un système de fichiers sur un niveau de stockage hautes performances. Dans d'autres cas, les modifications peuvent être uniquement pour des raisons d'administrateur, telles que l'isolation des fichiers de données d'un PDB donné dans des volumes dédiés.
3. Dans cet exemple, le `DATAFILE` la strophe reste inchangée, mais les journaux de reprise sont déplacés vers un nouvel emplacement dans `/redo` plutôt que de partager de l'espace avec les journaux d'archivage `/logs`.

```
CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
    MAXLOGFILES 16
    MAXLOGMEMBERS 3
    MAXDATAFILES 100
    MAXINSTANCES 8
    MAXLOGHISTORY 292
LOGFILE
  GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
  GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
-- STANDBY LOGFILE
DATAFILE
  '/oradata/WAFFLE/system01.dbf',
  '/oradata/WAFFLE/sysaux01.dbf',
  '/oradata/WAFFLE/undotbs01.dbf',
  '/oradata/WAFFLE/users01.dbf'
CHARACTER SET WE8MSWIN1252
;
```

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
SQL> CREATE CONTROLFILE REUSE DATABASE "WAFFLE" NORESETLOGS  ARCHIVELOG
  2      MAXLOGFILES 16
  3      MAXLOGMEMBERS 3
  4      MAXDATAFILES 100
  5      MAXINSTANCES 8
  6      MAXLOGHISTORY 292
  7  LOGFILE
  8    GROUP 1 '/redo/redo01.log'  SIZE 50M BLOCKSIZE 512,
  9    GROUP 2 '/redo/redo02.log'  SIZE 50M BLOCKSIZE 512,
 10    GROUP 3 '/redo/redo03.log'  SIZE 50M BLOCKSIZE 512
 11  -- STANDBY LOGFILE
 12  DATAFILE
 13    '/oradata/WAFFLE/system01.dbf',
 14    '/oradata/WAFFLE/sysaux01.dbf',
 15    '/oradata/WAFFLE/undotbs01.dbf',
 16    '/oradata/WAFFLE/users01.dbf'
 17  CHARACTER SET WE8MSWIN1252
 18  ;
Control file created.
SQL>
```

Si des fichiers sont mal placés ou si des paramètres sont mal configurés, des erreurs sont générées et indiquent ce qui doit être corrigé. La base de données est montée, mais elle n’est pas encore ouverte et ne peut pas être ouverte car les fichiers de données utilisés sont toujours marqués comme étant en mode de sauvegarde à chaud. Les journaux d’archivage doivent d’abord être appliqués pour rendre la base de données cohérente.

Réplication initiale du journal

Au moins une opération de réponse de journal est nécessaire pour rendre les fichiers de données cohérents. De nombreuses options sont disponibles pour relire les journaux. Dans certains cas, l’emplacement du journal d’archivage d’origine sur le serveur d’origine peut être partagé via NFS et la réponse du journal peut être effectuée directement. Dans d’autres cas, les journaux d’archivage doivent être copiés.

Par exemple, un simple `scp` l’opération peut copier tous les journaux en cours du serveur source vers le serveur de migration :

```
[oracle@jfsc3 arch]$ scp jfsc1:/logs/WAFFLE/arch/* ./
oracle@jfsc1's password:
1_22_912662036.dbf                                          100%   47MB
47.0MB/s   00:01
1_23_912662036.dbf                                          100%   40MB
40.4MB/s   00:00
1_24_912662036.dbf                                          100%   45MB
45.4MB/s   00:00
1_25_912662036.dbf                                          100%   41MB
40.9MB/s   00:01
1_26_912662036.dbf                                          100%   39MB
39.4MB/s   00:00
1_27_912662036.dbf                                          100%   39MB
38.7MB/s   00:00
1_28_912662036.dbf                                          100%   40MB
40.1MB/s   00:01
1_29_912662036.dbf                                          100%   17MB
16.9MB/s   00:00
1_30_912662036.dbf                                          100%  636KB
636.0KB/s   00:00
```

Relecture initiale du journal

Une fois les fichiers à l'emplacement du journal d'archivage, ils peuvent être relus en exécutant la commande `recover database until cancel` suivi de la réponse `AUTO` pour relire automatiquement tous les journaux disponibles.

```
SQL> recover database until cancel;
ORA-00279: change 382713 generated at 05/24/2016 09:00:54 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_23_912662036.dbf
ORA-00280: change 382713 for thread 1 is in sequence #23
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 405712 generated at 05/24/2016 15:01:05 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_24_912662036.dbf
ORA-00280: change 405712 for thread 1 is in sequence #24
ORA-00278: log file '/logs/WAFFLE/arch/1_23_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
ORA-00278: log file '/logs/WAFFLE/arch/1_30_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_31_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

La réponse finale au journal d'archivage signale une erreur, mais c'est normal. Le journal l'indique `sqlplus` a cherché un fichier journal particulier et ne l'a pas trouvé. La raison est, très probablement, que le fichier journal n'existe pas encore.

Si la base de données source peut être arrêtée avant de copier les journaux d'archivage, cette étape ne doit être effectuée qu'une seule fois. Les journaux d'archivage sont copiés et relus. Le processus peut ensuite se poursuivre directement vers le processus de mise en service qui réplique les journaux de reprise critiques.

Réplication et relecture incrémentielles du journal

Dans la plupart des cas, la migration n'est pas effectuée immédiatement. La fin du processus de migration peut prendre plusieurs jours, voire plusieurs semaines, ce qui signifie que les journaux doivent être envoyés en continu à la base de données de réplica et relus. Par conséquent, lors de la mise en service, un nombre minimal de données doit être transféré et relu.

Cela peut être scripté de plusieurs manières, mais l'une des méthodes les plus courantes est l'utilisation de rsync, un utilitaire commun de réplication de fichiers. La façon la plus sûre d'utiliser cet utilitaire est de le configurer en tant que démon. Par exemple, le `rsyncd.conf` le fichier suivant montre comment créer une ressource appelée `waffle.arch` Accessible avec les informations d'identification d'utilisateur Oracle et mappé sur `/logs/WAFFLE/arch`. Plus important encore, la ressource est définie en lecture seule, ce qui permet de lire les données de production sans les modifier.

```
[root@jfsc1 arch]# cat /etc/rsyncd.conf
[waffle.arch]
   uid=oracle
   gid=dba
   path=/logs/WAFFLE/arch
   read only = true
[root@jfsc1 arch]# rsync --daemon
```

La commande suivante synchronise la destination du journal d'archive du nouveau serveur avec la ressource rsync `waffle.arch` sur le serveur d'origine. Le `t` argument dans `rsync - potg` permet de comparer la liste de fichiers en fonction de l'horodatage et de copier uniquement les nouveaux fichiers. Ce processus fournit une mise à jour incrémentielle du nouveau serveur. Cette commande peut également être planifiée en cron pour s'exécuter de façon régulière.

```
[oracle@jfsc3 arch]$ rsync -potg --stats --progress jfsc1::waffle.arch/*
/logs/WAFFLE/arch/
1_31_912662036.dbf
      650240 100%  124.02MB/s    0:00:00 (xfer#1, to-check=8/18)
1_32_912662036.dbf
     4873728 100%  110.67MB/s    0:00:00 (xfer#2, to-check=7/18)
1_33_912662036.dbf
     4088832 100%   50.64MB/s    0:00:00 (xfer#3, to-check=6/18)
1_34_912662036.dbf
     8196096 100%   54.66MB/s    0:00:00 (xfer#4, to-check=5/18)
1_35_912662036.dbf
    19376128 100%   57.75MB/s    0:00:00 (xfer#5, to-check=4/18)
1_36_912662036.dbf
       71680 100%  201.15kB/s    0:00:00 (xfer#6, to-check=3/18)
1_37_912662036.dbf
     1144320 100%    3.06MB/s    0:00:00 (xfer#7, to-check=2/18)
1_38_912662036.dbf
    35757568 100%   63.74MB/s    0:00:00 (xfer#8, to-check=1/18)
1_39_912662036.dbf
      984576 100%    1.63MB/s    0:00:00 (xfer#9, to-check=0/18)
Number of files: 18
Number of files transferred: 9
Total file size: 399653376 bytes
Total transferred file size: 75143168 bytes
Literal data: 75143168 bytes
Matched data: 0 bytes
File list size: 474
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 204
Total bytes received: 75153219
sent 204 bytes  received 75153219 bytes  150306846.00 bytes/sec
total size is 399653376  speedup is 5.32
```

Une fois les journaux reçus, ils doivent être relus. Les exemples précédents montrent l'utilisation de sqlplus pour une exécution manuelle `recover database until cancel`, un processus qui peut être facilement automatisé. L'exemple illustré ici utilise le script décrit dans "Relire les journaux sur la base de données". Les scripts acceptent un argument qui spécifie la base de données nécessitant une opération de relecture. Cela permet d'utiliser le même script dans un effort de migration multibase de données.

```
[oracle@jfsc3 logs]$ ./replay.logs.pl WAFFLE
ORACLE_SID = [WAFFLE] ? The Oracle base remains unchanged with value
/orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu May 26 10:47:16 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 713874 generated at 05/26/2016 04:26:43 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_31_912662036.dbf
ORA-00280: change 713874 for thread 1 is in sequence #31
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 814256 generated at 05/26/2016 04:52:30 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_32_912662036.dbf
ORA-00280: change 814256 for thread 1 is in sequence #32
ORA-00278: log file '/logs/WAFFLE/arch/1_31_912662036.dbf' no longer
needed for
this recovery
ORA-00279: change 814780 generated at 05/26/2016 04:53:04 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_33_912662036.dbf
ORA-00280: change 814780 for thread 1 is in sequence #33
ORA-00278: log file '/logs/WAFFLE/arch/1_32_912662036.dbf' no longer
needed for
this recovery
...
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
ORA-00278: log file '/logs/WAFFLE/arch/1_39_912662036.dbf' no longer
needed for
this recovery
ORA-00308: cannot open archived log '/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

Mise en service

Lorsque vous êtes prêt à passer au nouvel environnement, vous devez effectuer une synchronisation finale qui inclut à la fois les journaux d'archivage et les journaux de reprise. Si l'emplacement original du journal de reprise n'est pas déjà connu, il peut être identifié comme suit :

```
SQL> select member from v$logfile;
MEMBER
------------------------------------------------------------------------
------
/logs/WAFFLE/redo/redo01.log
/logs/WAFFLE/redo/redo02.log
/logs/WAFFLE/redo/redo03.log
```

1. Arrêtez la base de données source.
2. Effectuez une synchronisation finale des journaux d'archivage sur le nouveau serveur avec la méthode souhaitée.
3. Les fichiers redo log source doivent être copiés sur le nouveau serveur. Dans cet exemple, les journaux de reprise ont été déplacés vers un nouveau répertoire à `/redo`.

   ```
   [oracle@jfsc3 logs]$ scp jfsc1:/logs/WAFFLE/redo/* /redo/
   oracle@jfsc1's password:
   redo01.log
   100%   50MB  50.0MB/s   00:01
   redo02.log
   100%   50MB  50.0MB/s   00:00
   redo03.log
   100%   50MB  50.0MB/s   00:00
   ```

4. À ce stade, le nouvel environnement de base de données contient tous les fichiers nécessaires pour le ramener au même état que la source. Les journaux d'archivage doivent être relus une dernière fois.

```
SQL> recover database until cancel;
ORA-00279: change 1120099 generated at 05/26/2016 09:59:21 needed for
thread 1
ORA-00289: suggestion : /logs/WAFFLE/arch/1_40_912662036.dbf
ORA-00280: change 1120099 for thread 1 is in sequence #40
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
ORA-00308: cannot open archived log
'/logs/WAFFLE/arch/1_40_912662036.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

5. Une fois l'opération terminée, les journaux de reprise doivent être relus. Si le message s'affiche `Media recovery complete` est renvoyé, le processus a réussi et les bases de données sont synchronisées et peuvent être ouvertes.

```
SQL> recover database;
Media recovery complete.
SQL> alter database open;
Database altered.
```

Envoi de journaux - ASM vers le système de fichiers

Cet exemple illustre l'utilisation d'Oracle RMAN pour migrer une base de données. Il est très similaire à l'exemple précédent de système de fichiers pour l'envoi de journaux de système de fichiers, mais les fichiers sur ASM ne sont pas visibles par l'hôte. Les seules options de migration des données situées sur les périphériques ASM sont soit le déplacement du LUN ASM, soit l'utilisation d'Oracle RMAN pour effectuer les opérations de copie.

Bien que RMAN soit obligatoire pour la copie de fichiers à partir d'Oracle ASM, l'utilisation de RMAN ne se limite pas à ASM. RMAN peut être utilisé pour migrer de tout type de stockage vers tout autre type.

Cet exemple montre le déplacement d'une base de données appelée PANCAKE depuis le stockage ASM vers un système de fichiers standard situé sur un serveur différent au niveau des chemins `/oradata` et `/logs`.

Créer une sauvegarde de base de données

La première étape consiste à créer une sauvegarde de la base de données à migrer vers un autre serveur. Comme la source utilise Oracle ASM, RMAN doit être utilisé. Une simple sauvegarde RMAN peut être effectuée comme suit. Cette méthode crée une sauvegarde balisée qui peut être facilement identifiée par RMAN plus tard dans la procédure.

La première commande définit le type de destination de la sauvegarde et l'emplacement à utiliser. La seconde lance la sauvegarde des fichiers de données uniquement.

```
RMAN> configure channel device type disk format '/rman/pancake/%U';
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT   '/rman/pancake/%U';
new RMAN configuration parameters are successfully stored
RMAN> backup database tag 'ONTAP_MIGRATION';
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=251 device type=DISK
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/PANCAKE/system01.dbf
input datafile file number=00002 name=+ASM0/PANCAKE/sysaux01.dbf
input datafile file number=00003 name=+ASM0/PANCAKE/undotbs101.dbf
input datafile file number=00004 name=+ASM0/PANCAKE/users01.dbf
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1gr6c161_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:03
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/1hr6c164_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

Fichier de contrôle de sauvegarde

Un fichier de contrôle de sauvegarde est requis plus tard dans la procédure pour `duplicate database` fonctionnement.

```
RMAN> backup current controlfile format '/rman/pancake/ctrl.bkp';
Starting backup at 24-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting full datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
channel ORA_DISK_1: starting piece 1 at 24-MAY-16
channel ORA_DISK_1: finished piece 1 at 24-MAY-16
piece handle=/rman/pancake/ctrl.bkp tag=TAG20160524T032651 comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

Sauvegarde du fichier de paramètres

Un fichier de paramètres est également requis dans le nouvel environnement. La méthode la plus simple consiste à créer un fichier pfile à partir du fichier spfile ou pfile actuel. Dans cet exemple, la base de données source utilise un fichier spfile.

```
RMAN> create pfile='/rman/pancake/pfile' from spfile;
Statement processed
```

Script de renommage de fichier ASM

Plusieurs emplacements de fichiers actuellement définis dans les fichiers de contrôle changent lorsque la base de données est déplacée. Le script suivant crée un script RMAN pour faciliter le processus. Cet exemple illustre une base de données comportant un très petit nombre de fichiers de données, mais en général, les bases de données contiennent des centaines, voire des milliers de fichiers de données.

Ce script est disponible dans "Conversion de noms de système de fichiers ASM en système de fichiers" et il fait deux choses.

Tout d'abord, il crée un paramètre pour redéfinir les emplacements du journal de reprise appelés `log_file_name_convert`. Il s'agit essentiellement d'une liste de champs alternatifs. Le premier champ est l'emplacement d'un journal de reprise en cours et le second est l'emplacement sur le nouveau serveur. Le schéma est alors répété.

La deuxième fonction consiste à fournir un modèle pour renommer le fichier de données. Le script passe en boucle dans les fichiers de données, extrait les informations relatives au nom et au numéro de fichier et les formate en tant que script RMAN. Il fait ensuite la même chose avec les fichiers temporaires. Le résultat est un script rman simple qui peut être modifié comme vous le souhaitez pour vous assurer que les fichiers sont restaurés à l'emplacement souhaité.

```
SQL> @/rman/mk.rename.scripts.sql
Parameters for log file conversion:
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/NEW_PATH/redo01.log','+ASM0/PANCAKE/redo02.log',
'/NEW_PATH/redo02.log','+ASM0/PANCAKE/redo03.log', '/NEW_PATH/redo03.log'
rman duplication script:
run
{
set newname for datafile 1 to '+ASM0/PANCAKE/system01.dbf';
set newname for datafile 2 to '+ASM0/PANCAKE/sysaux01.dbf';
set newname for datafile 3 to '+ASM0/PANCAKE/undotbs101.dbf';
set newname for datafile 4 to '+ASM0/PANCAKE/users01.dbf';
set newname for tempfile 1 to '+ASM0/PANCAKE/temp01.dbf';
duplicate target database for standby backup location INSERT_PATH_HERE;
}
PL/SQL procedure successfully completed.
```

Capturer la sortie de cet écran. Le `log_file_name_convert` le paramètre est placé dans le fichier pfile comme décrit ci-dessous. Le script de renommage et de duplication du fichier de données RMAN doit être modifié en conséquence pour placer les fichiers de données aux emplacements souhaités. Dans cet exemple, ils sont tous placés dans `/oradata/pancake`.

```
run
{
set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
set newname for datafile 4 to '/oradata/pancake/users.dbf';
set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
duplicate target database for standby backup location '/rman/pancake';
}
```

Préparer la structure du répertoire

Les scripts sont presque prêts à être exécutés, mais d'abord la structure de répertoire doit être en place. Si les répertoires requis ne sont pas déjà présents, ils doivent être créés ou la procédure de démarrage de la base de données échoue. L'exemple ci-dessous reflète les exigences minimales.

```
[oracle@jfsc2 ~]$ mkdir /oradata/pancake
[oracle@jfsc2 ~]$ mkdir /logs/pancake
[oracle@jfsc2 ~]$ cd /orabin/admin
[oracle@jfsc2 admin]$ mkdir PANCAKE
[oracle@jfsc2 admin]$ cd PANCAKE
[oracle@jfsc2 PANCAKE]$ mkdir adump dpdump pfile scripts xdb_wallet
```

Créer une entrée oratab

La commande suivante est requise pour que des utilitaires tels que oraenv fonctionnent correctement.

```
PANCAKE:/orabin/product/12.1.0/dbhome_1:N
```

Mises à jour des paramètres

Le fichier pfile enregistré doit être mis à jour pour refléter toute modification de chemin sur le nouveau serveur. Les modifications du chemin d'accès au fichier de données sont modifiées par le script de duplication RMAN, et presque toutes les bases de données nécessitent des modifications `control_files` et `log_archive_dest` paramètres. Il peut également y avoir des emplacements de fichiers d'audit qui doivent être modifiés, ainsi que des paramètres tels que `db_create_file_dest` Peut ne pas être pertinent en dehors d'ASM. Un administrateur de base de données expérimenté doit examiner attentivement les modifications proposées avant de poursuivre.

Dans cet exemple, les changements de clé sont les emplacements des fichiers de contrôle, la destination de l'archive de journal et l'ajout du `log_file_name_convert` paramètre.

```
PANCAKE.__data_transfer_cache_size=0
PANCAKE.__db_cache_size=545259520
PANCAKE.__java_pool_size=4194304
PANCAKE.__large_pool_size=25165824
PANCAKE.__oracle_base='/orabin'#ORACLE_BASE set from environment
PANCAKE.__pga_aggregate_target=268435456
PANCAKE.__sga_target=805306368
PANCAKE.__shared_io_pool_size=29360128
PANCAKE.__shared_pool_size=192937984
PANCAKE.__streams_pool_size=0
*.audit_file_dest='/orabin/admin/PANCAKE/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='+ASM0/PANCAKE/control01.ctl','+ASM0/PANCAKE/control02.ctl
'
*.control_files='/oradata/pancake/control01.ctl','/logs/pancake/control02.
ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='PANCAKE'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=PANCAKEXDB)'
*.log_archive_dest_1='LOCATION=+ASM1'
*.log_archive_dest_1='LOCATION=/logs/pancake'
*.log_archive_format='%t_%s_%r.dbf'
'/logs/path/redo02.log'
*.log_file_name_convert = '+ASM0/PANCAKE/redo01.log',
'/logs/pancake/redo01.log', '+ASM0/PANCAKE/redo02.log',
'/logs/pancake/redo02.log', '+ASM0/PANCAKE/redo03.log',
'/logs/pancake/redo03.log'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

Une fois les nouveaux paramètres confirmés, les paramètres doivent être mis en vigueur. Plusieurs options existent, mais la plupart des clients créent un fichier spfile basé sur le fichier pfile texte.

```
bash-4.1$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0 Production on Fri Jan 8 11:17:40 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> create spfile from pfile='/rman/pancake/pfile';
File created.
```

Nom de démarrage

La dernière étape avant la réplication de la base de données consiste à afficher les processus de la base de données, mais pas à monter les fichiers. Dans cette étape, des problèmes avec le fichier spfile peuvent devenir évidents. Si le `startup nomount` la commande échoue en raison d'une erreur de paramètre, il est simple de s'arrêter, de corriger le modèle pfile, de le recharger en tant que fichier spfile et de réessayer.

```
SQL> startup nomount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             373296240 bytes
Database Buffers          423624704 bytes
Redo Buffers                5455872 bytes
```

Dupliquez la base de données

La restauration de la sauvegarde RMAN précédente vers le nouvel emplacement prend plus de temps que les autres étapes de ce processus. La base de données doit être dupliquée sans modification de l'ID de base de données (DBID) ou réinitialisation des journaux. Cela empêche l'application des journaux, ce qui est une étape nécessaire pour synchroniser complètement les copies.

Connectez-vous à la base de données avec RMAN en tant qu'aux et exécutez la commande duplicate database en utilisant le script créé lors d'une étape précédente.

```
[oracle@jfsc2 pancake]$ rman auxiliary /
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 03:04:56
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to auxiliary database: PANCAKE (not mounted)
RMAN> run
2> {
3> set newname for datafile 1 to '/oradata/pancake/pancake.dbf';
4> set newname for datafile 2 to '/oradata/pancake/sysaux.dbf';
5> set newname for datafile 3 to '/oradata/pancake/undotbs1.dbf';
6> set newname for datafile 4 to '/oradata/pancake/users.dbf';
7> set newname for tempfile 1 to '/oradata/pancake/temp.dbf';
```

```
8> duplicate target database for standby backup location '/rman/pancake';
9> }
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting Duplicate Db at 24-MAY-16
contents of Memory Script:
{
   restore clone standby controlfile from  '/rman/pancake/ctrl.bkp';
}
executing Memory Script
Starting restore at 24-MAY-16
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
channel ORA_AUX_DISK_1: restoring control file
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:01
output file name=/oradata/pancake/control01.ctl
output file name=/logs/pancake/control02.ctl
Finished restore at 24-MAY-16
contents of Memory Script:
{
   sql clone 'alter database mount standby database';
}
executing Memory Script
sql statement: alter database mount standby database
released channel: ORA_AUX_DISK_1
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=243 device type=DISK
contents of Memory Script:
{
   set newname for tempfile  1 to
 "/oradata/pancake/temp.dbf";
   switch clone tempfile all;
   set newname for datafile  1 to
 "/oradata/pancake/pancake.dbf";
   set newname for datafile  2 to
 "/oradata/pancake/sysaux.dbf";
   set newname for datafile  3 to
 "/oradata/pancake/undotbs1.dbf";
   set newname for datafile  4 to
 "/oradata/pancake/users.dbf";
   restore
   clone database
   ;
```

```
}
executing Memory Script
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/pancake/temp.dbf in control file
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
executing command: SET NEWNAME
Starting restore at 24-MAY-16
using channel ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: starting datafile backup set restore
channel ORA_AUX_DISK_1: specifying datafile(s) to restore from backup set
channel ORA_AUX_DISK_1: restoring datafile 00001 to
/oradata/pancake/pancake.dbf
channel ORA_AUX_DISK_1: restoring datafile 00002 to
/oradata/pancake/sysaux.dbf
channel ORA_AUX_DISK_1: restoring datafile 00003 to
/oradata/pancake/undotbs1.dbf
channel ORA_AUX_DISK_1: restoring datafile 00004 to
/oradata/pancake/users.dbf
channel ORA_AUX_DISK_1: reading from backup piece
/rman/pancake/1gr6c161_1_1
channel ORA_AUX_DISK_1: piece handle=/rman/pancake/1gr6c161_1_1
tag=ONTAP_MIGRATION
channel ORA_AUX_DISK_1: restored backup piece 1
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:07
Finished restore at 24-MAY-16
contents of Memory Script:
{
   switch clone datafile all;
}
executing Memory Script
datafile 1 switched to datafile copy
input datafile copy RECID=5 STAMP=912655725 file
name=/oradata/pancake/pancake.dbf
datafile 2 switched to datafile copy
input datafile copy RECID=6 STAMP=912655725 file
name=/oradata/pancake/sysaux.dbf
datafile 3 switched to datafile copy
input datafile copy RECID=7 STAMP=912655725 file
name=/oradata/pancake/undotbs1.dbf
datafile 4 switched to datafile copy
input datafile copy RECID=8 STAMP=912655725 file
name=/oradata/pancake/users.dbf
Finished Duplicate Db at 24-MAY-16
```

Réplication initiale du journal

Vous devez maintenant envoyer les modifications de la base de données source vers un nouvel emplacement. Cela peut nécessiter une combinaison d'étapes. La méthode la plus simple serait que RMAN sur la base de données source écrive des journaux d'archive sur une connexion réseau partagée. Si aucun emplacement partagé n'est disponible, une autre méthode consiste à utiliser RMAN pour écrire dans un système de fichiers local, puis à utiliser rcp ou rsync pour copier les fichiers.

Dans cet exemple, le `/rman` Directory est un partage NFS disponible pour la base de données d'origine et migrée.

L'une des questions importantes est la `disk format` clause. Le format de disque de la sauvegarde est `%h_%e_%a.dbf`, Ce qui signifie que vous devez utiliser le format du numéro de thread, du numéro de séquence et de l'ID d'activation de la base de données. Bien que les lettres soient différentes, cela correspond à `log_archive_format='%t_%s_%r.dbf` dans le fichier pfile. Ce paramètre spécifie également les journaux d'archivage au format de numéro de thread, de numéro de séquence et d'ID d'activation. Le résultat final est que les sauvegardes du fichier journal sur la source utilisent une convention de dénomination attendue par la base de données. Cela permet de réaliser des opérations telles que `recover database` beaucoup plus simple parce que sqlplus anticipe correctement les noms des journaux d'archive à lire.

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/arch/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
released channel: ORA_DISK_1
RMAN> backup as copy archivelog from time 'sysdate-2';
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=373 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=70 STAMP=912658508
output file name=/rman/pancake/logship/1_54_912576125.dbf RECID=123
STAMP=912659482
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=29 STAMP=912654101
output file name=/rman/pancake/logship/1_41_912576125.dbf RECID=124
STAMP=912659483
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=33 STAMP=912654688
output file name=/rman/pancake/logship/1_45_912576125.dbf RECID=152
STAMP=912659514
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=36 STAMP=912654809
output file name=/rman/pancake/logship/1_47_912576125.dbf RECID=153
STAMP=912659515
channel ORA_DISK_1: archived log copy complete, elapsed time: 00:00:01
Finished backup at 24-MAY-16
```

Relecture initiale du journal

Une fois les fichiers à l'emplacement du journal d'archivage, ils peuvent être relus en exécutant la commande `recover database until cancel` suivi de la réponse `AUTO` pour relire automatiquement tous les journaux disponibles. Le fichier de paramètres dirige actuellement les journaux d'archivage vers `/logs/archive`, Mais cela ne correspond pas à l'emplacement où RMAN a été utilisé pour enregistrer les journaux. L'emplacement peut être redirigé temporairement comme suit avant de récupérer la base de données.

```
SQL> alter system set log_archive_dest_1='LOCATION=/rman/pancake/logship'
scope=memory;
System altered.
SQL> recover standby database until cancel;
ORA-00279: change 560224 generated at 05/24/2016 03:25:53 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_49_912576125.dbf
ORA-00280: change 560224 for thread 1 is in sequence #49
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
AUTO
ORA-00279: change 560353 generated at 05/24/2016 03:29:17 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_50_912576125.dbf
ORA-00280: change 560353 for thread 1 is in sequence #50
ORA-00278: log file '/rman/pancake/logship/1_49_912576125.dbf' no longer
needed
for this recovery
...
ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
ORA-00278: log file '/rman/pancake/logship/1_53_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_54_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
```

La réponse finale au journal d'archivage signale une erreur, mais c'est normal. L'erreur indique que sqlplus recherchait un fichier journal particulier et qu'il ne l'a pas trouvé. La raison est la plus probable que le fichier journal n'existe pas encore.

Si la base de données source peut être arrêtée avant de copier les journaux d'archivage, cette étape ne doit être effectuée qu'une seule fois. Les journaux d'archivage sont copiés et relus. Le processus peut ensuite se poursuivre directement vers le processus de mise en service qui réplique les journaux de reprise critiques.

Réplication et relecture incrémentielles du journal

Dans la plupart des cas, la migration n'est pas effectuée immédiatement. La fin du processus de migration peut prendre plusieurs jours, voire plusieurs semaines, ce qui signifie que les journaux doivent être envoyés en continu à la base de données de réplica et relus. Ainsi, le transfert et la lecture de données minimales doivent être assurés à l'arrivée de la mise en service.

Ce processus peut facilement être scripté. Par exemple, la commande suivante peut être planifiée sur la base de données d'origine pour s'assurer que l'emplacement utilisé pour l'envoi des journaux est mis à jour en

permanence.

```
[oracle@jfsc1 pancake]$ cat copylogs.rman
configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
backup as copy archivelog from time 'sysdate-2';
```

```
[oracle@jfsc1 pancake]$ rman target / cmdfile=copylogs.rman
Recovery Manager: Release 12.1.0.2.0 - Production on Tue May 24 04:36:19
2016
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: PANCAKE (DBID=3574534589)
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
current log archived
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=369 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:22
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=41 RECID=124 STAMP=912659483
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:23
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_41_912576125.dbf
continuing other job steps, job failed will not be re-run
...
```

```
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:55
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at 05/24/2016
04:36:57
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
Recovery Manager complete.
```

Une fois les journaux reçus, ils doivent être relus. Des exemples précédents ont montré l'utilisation de sqlplus pour une exécution manuelle `recover database until cancel`, qui peut être facilement automatisé. L'exemple illustré ici utilise le script décrit dans "Relire les journaux sur la base de données de secours". Le script accepte un argument qui spécifie la base de données nécessitant une opération de relecture. Ce processus permet d'utiliser le même script dans un effort de migration multibase de données.

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 04:47:10 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 560591 generated at 05/24/2016 03:33:56 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_54_912576125.dbf
ORA-00280: change 560591 for thread 1 is in sequence #54
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 562219 generated at 05/24/2016 04:15:08 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_55_912576125.dbf
ORA-00280: change 562219 for thread 1 is in sequence #55
ORA-00278: log file '/rman/pancake/logship/1_54_912576125.dbf' no longer
needed for this recovery
ORA-00279: change 562370 generated at 05/24/2016 04:19:18 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_56_912576125.dbf
ORA-00280: change 562370 for thread 1 is in sequence #56
ORA-00278: log file '/rman/pancake/logship/1_55_912576125.dbf' no longer
needed for this recovery
...
ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
ORA-00278: log file '/rman/pancake/logship/1_64_912576125.dbf' no longer
needed for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_65_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

Mise en service

Lorsque vous êtes prêt à passer au nouvel environnement, vous devez effectuer une synchronisation finale. Lorsque vous travaillez avec des systèmes de fichiers réguliers, il est facile de s'assurer que la base de données migrée est synchronisée à 100 % par rapport à l'original car les journaux de reprise d'origine sont copiés et relus. Il n'y a pas de bonne façon de le faire avec ASM. Seuls les journaux d'archivage peuvent être facilement recopiés. Pour s'assurer qu'aucune donnée n'est perdue, l'arrêt final de la base de données d'origine doit être effectué avec précaution.

1. Tout d'abord, la base de données doit être mise en veille, en veillant à ce qu'aucune modification ne soit apportée. Cette mise en veille peut inclure la désactivation des opérations planifiées, l'arrêt des auditeurs et/ou l'arrêt des applications.
2. Une fois cette étape effectuée, la plupart des administrateurs de bases de données créent une table fictive qui sert de marqueur de l'arrêt.
3. Forcer l'archivage des journaux pour s'assurer que la création de la table fictive est enregistrée dans les journaux d'archivage. Pour ce faire, exécutez les commandes suivantes :

```
SQL> create table cutovercheck as select * from dba_users;
Table created.
SQL> alter system archive log current;
System altered.
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
```

4. Pour copier le dernier des journaux d'archivage, exécutez les commandes suivantes. La base de données doit être disponible mais pas ouverte.

```
SQL> startup mount;
ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             331353200 bytes
Database Buffers          465567744 bytes
Redo Buffers                5455872 bytes
Database mounted.
```

5. Pour copier les journaux d'archivage, exécutez les commandes suivantes :

```
RMAN> configure channel device type disk format
'/rman/pancake/logship/%h_%e_%a.dbf';
2> backup as copy archivelog from time 'sysdate-2';
3>
4>
using target database control file instead of recovery catalog
old RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters:
CONFIGURE CHANNEL DEVICE TYPE DISK FORMAT
'/rman/pancake/logship/%h_%e_%a.dbf';
new RMAN configuration parameters are successfully stored
Starting backup at 24-MAY-16
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=8 device type=DISK
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=54 RECID=123 STAMP=912659482
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:24
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_54_912576125.dbf
continuing other job steps, job failed will not be re-run
...
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=45 RECID=152 STAMP=912659514
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:58:58
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_45_912576125.dbf
continuing other job steps, job failed will not be re-run
channel ORA_DISK_1: starting archived log copy
input archived log thread=1 sequence=47 RECID=153 STAMP=912659515
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03009: failure of backup command on ORA_DISK_1 channel at
05/24/2016 04:59:00
ORA-19635: input and output file names are identical:
/rman/pancake/logship/1_47_912576125.dbf
```

6. Enfin, rejouez les journaux d’archive restants sur le nouveau serveur.

```
[root@jfsc2 pancake]# ./replaylogs.pl PANCAKE
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Tue May 24 05:00:53 2016
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> ORA-00279: change 563137 generated at 05/24/2016 04:36:20 needed
for thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_65_912576125.dbf
ORA-00280: change 563137 for thread 1 is in sequence #65
Specify log: {<RET>=suggested | filename | AUTO | CANCEL}
ORA-00279: change 563629 generated at 05/24/2016 04:55:20 needed for
thread 1
ORA-00289: suggestion : /rman/pancake/logship/1_66_912576125.dbf
ORA-00280: change 563629 for thread 1 is in sequence #66
ORA-00278: log file '/rman/pancake/logship/1_65_912576125.dbf' no longer
needed
for this recovery
ORA-00308: cannot open archived log
'/rman/pancake/logship/1_66_912576125.dbf'
ORA-27037: unable to obtain file status
Linux-x86_64 Error: 2: No such file or directory
Additional information: 3
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
```

7. À ce stade, répliquez toutes les données. La base de données est prête à être convertie à partir d'une base de données de secours vers une base de données opérationnelle active, puis ouverte.

```
SQL> alter database activate standby database;
Database altered.
SQL> alter database open;
Database altered.
```

8. Confirmer la présence de la table factice, puis la déposer.

```
SQL> desc cutovercheck
 Name                                      Null?    Type
 ----------------------------------------- --------
----------------------------
 USERNAME                                  NOT NULL VARCHAR2(128)
 USER_ID                                   NOT NULL NUMBER
 PASSWORD                                           VARCHAR2(4000)
 ACCOUNT_STATUS                            NOT NULL VARCHAR2(32)
 LOCK_DATE                                          DATE
 EXPIRY_DATE                                        DATE
 DEFAULT_TABLESPACE                        NOT NULL VARCHAR2(30)
 TEMPORARY_TABLESPACE                      NOT NULL VARCHAR2(30)
 CREATED                                   NOT NULL DATE
 PROFILE                                   NOT NULL VARCHAR2(128)
 INITIAL_RSRC_CONSUMER_GROUP                        VARCHAR2(128)
 EXTERNAL_NAME                                      VARCHAR2(4000)
 PASSWORD_VERSIONS                                  VARCHAR2(12)
 EDITIONS_ENABLED                                   VARCHAR2(1)
 AUTHENTICATION_TYPE                                VARCHAR2(8)
 PROXY_ONLY_CONNECT                                 VARCHAR2(1)
 COMMON                                             VARCHAR2(3)
 LAST_LOGIN                                         TIMESTAMP(9) WITH
TIME ZONE
 ORACLE_MAINTAINED                                  VARCHAR2(1)
SQL> drop table cutovercheck;
Table dropped.
```

Migration des journaux de reprise sans interruption

Il arrive qu'une base de données soit correctement organisée de manière globale, à l'exception des journaux de reprise. Cela peut se produire pour de nombreuses raisons, dont la plus courante est liée aux snapshots. Des produits tels que SnapManager pour Oracle, SnapCenter et la structure de gestion du stockage NetApp Snap Creator permettent une restauration quasi instantanée d'une base de données, mais uniquement si vous restaurez l'état des volumes de fichiers de données. Si les journaux de reprise partagent l'espace avec les fichiers de données, la restauration ne peut pas être effectuée en toute sécurité, car elle entraînerait la destruction des journaux de reprise, ce qui entraînerait probablement une perte des données. Les journaux de reprise doivent donc être déplacés.

Cette procédure est simple et peut être effectuée sans interruption.

Configuration actuelle du journal de reprise

1. Identifiez le nombre de groupes de fichiers redo log et leurs numéros de groupe respectifs.

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
--------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
rows selected.
```

2. Indiquez la taille des journaux de reprise.

```
SQL> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
--------------------------------------------------------------------------
--------
1 524288000
2 524288000
3 524288000
```

Créer de nouveaux journaux

1. Pour chaque journal de reprise, créez un nouveau groupe avec la taille et le nombre de membres correspondants.

```
SQL> alter database add logfile ('/newredo0/redo01a.log',
'/newredo1/redo01b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo02a.log',
'/newredo1/redo02b.log') size 500M;
Database altered.
SQL> alter database add logfile ('/newredo0/redo03a.log',
'/newredo1/redo03b.log') size 500M;
Database altered.
SQL>
```

2. Vérifiez la nouvelle configuration.

```
SQL> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
---------------------------------------------------------------------------
--------
1 /redo0/NTAP/redo01a.log
1 /redo1/NTAP/redo01b.log
2 /redo0/NTAP/redo02a.log
2 /redo1/NTAP/redo02b.log
3 /redo0/NTAP/redo03a.log
3 /redo1/NTAP/redo03b.log
4 /newredo0/redo01a.log
4 /newredo1/redo01b.log
5 /newredo0/redo02a.log
5 /newredo1/redo02b.log
6 /newredo0/redo03a.log
6 /newredo1/redo03b.log
12 rows selected.
```

Supprimez les anciens journaux

1. Supprimez les anciens journaux (groupes 1, 2 et 3).

```
SQL> alter database drop logfile group 1;
Database altered.
SQL> alter database drop logfile group 2;
Database altered.
SQL> alter database drop logfile group 3;
Database altered.
```

2. Si vous rencontrez une erreur qui vous empêche de supprimer un journal actif, forcez un commutateur au journal suivant pour libérer le verrouillage et forcer un point de contrôle global. Reportez-vous à l'exemple suivant de ce processus. La tentative de suppression du groupe de fichiers journaux 2, qui se trouvait sur l'ancien emplacement, a été refusée parce qu'il y avait encore des données actives dans ce fichier journal.

```
SQL> alter database drop logfile group 2;
alter database drop logfile group 2
*
ERROR at line 1:
ORA-01623: log 2 is current log for instance NTAP (thread 1) - cannot
drop
ORA-00312: online log 2 thread 1: '/redo0/NTAP/redo02a.log'
ORA-00312: online log 2 thread 1: '/redo1/NTAP/redo02b.log'
```

3. Un archivage de journaux suivi d'un point de contrôle vous permet de supprimer le fichier journal.

```
SQL> alter system archive log current;
System altered.
SQL> alter system checkpoint;
System altered.
SQL> alter database drop logfile group 2;
Database altered.
```

4. Supprimez ensuite les journaux du système de fichiers. Vous devez effectuer ce processus avec une extrême prudence.

Copie de données hôte

À l'instar de la migration au niveau des bases de données, la migration au niveau de la couche hôte offre une approche indépendante du fournisseur de stockage.

En d'autres termes, parfois "juste copier les fichiers" est la meilleure option.

Bien que cette approche peu technologique puisse sembler trop basique, elle offre des avantages significatifs, car aucun logiciel spécial n'est requis et les données d'origine ne sont pas modifiées en toute sécurité pendant le processus. La principale limitation est le fait qu'une migration de données de copie de fichier est un processus perturbateur, car la base de données doit être arrêtée avant le début de l'opération de copie. Il n'y a pas de bonne façon de synchroniser les modifications dans un fichier, de sorte que les fichiers doivent être complètement suspendus avant le début de la copie.

Si l'arrêt requis par une opération de copie n'est pas souhaitable, la meilleure option basée sur l'hôte suivante consiste à exploiter un gestionnaire de volumes logiques (LVM). De nombreuses options LVM existent, y compris Oracle ASM, toutes avec des capacités similaires, mais avec certaines limitations qui doivent être prises en compte. Dans la plupart des cas, la migration peut s'effectuer sans interruption ni perturbation.

Copie du système de fichiers vers le système de fichiers

L'utilité d'une simple opération de copie ne doit pas être sous-estimée. Cette opération requiert un temps d'indisponibilité lors de la copie, mais le processus est extrêmement fiable et ne requiert aucune expertise particulière en matière de systèmes d'exploitation, de bases de données ou de systèmes de stockage. De plus, elle est très sûre car elle n'affecte pas les données d'origine. Généralement, un administrateur système modifie les systèmes de fichiers source pour qu'ils soient montés en lecture seule, puis redémarre un serveur pour garantir que rien ne risque d'endommager les données actuelles. Le processus de copie peut être scripté pour s'assurer qu'il s'exécute aussi rapidement que possible sans risque d'erreur de l'utilisateur. Comme le type d'E/S est un simple transfert séquentiel de données, il est très peu gourmand en bande passante.

L'exemple suivant illustre une option pour une migration sûre et rapide.

De production

L'environnement à migrer est le suivant :

- Systèmes de fichiers actuels

```
ontap-nfs1:/host1_oradata       52428800  16196928  36231872  31%
/oradata
ontap-nfs1:/host1_logs          49807360    548032  49259328   2% /logs
```

- Nouveaux systèmes de fichiers

```
ontap-nfs1:/host1_logs_new      49807360       128  49807232   1%
/new/logs
ontap-nfs1:/host1_oradata_new   49807360       128  49807232   1%
/new/oradata
```

Présentation

Il suffit à l'administrateur de bases de données de fermer la base de données et de copier les fichiers pour migrer la base de données. Toutefois, ce processus peut être facilement scripté si de nombreuses bases de données doivent être migrées ou si la réduction des temps d'indisponibilité est essentielle. L'utilisation de scripts réduit également les risques d'erreur de l'utilisateur.

Les exemples de scripts présentés automatisent les opérations suivantes :

- Arrêt de la base de données
- Conversion des systèmes de fichiers existants en état de lecture seule
- Copie de toutes les données de la source vers les systèmes de fichiers cibles, ce qui préserve toutes les autorisations de fichier
- Démontage de l'ancien et du nouveau système de fichiers
- Remontage des nouveaux systèmes de fichiers aux mêmes chemins que les systèmes de fichiers précédents

Procédure

1. Arrêtez la base de données.

```
[root@host1 current]# ./dbshut.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 15:58:48 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
```

2. Convertissez les systèmes de fichiers en lecture seule. Ceci peut être effectué plus rapidement en utilisant un script, comme indiqué dans la "Convertir le système de fichiers en lecture seule".

```
[root@host1 current]# ./mk.fs.readonly.pl /oradata
/oradata unmounted
/oradata mounted read-only
[root@host1 current]# ./mk.fs.readonly.pl /logs
/logs unmounted
/logs mounted read-only
```

3. Vérifiez que les systèmes de fichiers sont maintenant en lecture seule.

```
ontap-nfs1:/host1_oradata on /oradata type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_logs on /logs type nfs
(ro,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

4. Synchroniser le contenu du système de fichiers avec le `rsync` commande.

```
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /oradata/ /new/oradata/
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
```

```
 10737426432 100%  153.50MB/s    0:01:06 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%   12.09MB/s    0:00:01 (xfer#2, to-check=9/13)
...
NTAP/undotbs02.dbf
  1073750016 100%  131.60MB/s    0:00:07 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    3.95MB/s    0:00:01 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  162204017.96 bytes/sec
total size is 18570092218  speedup is 1.00
[root@host1 current]# rsync -rlpogt --stats --progress
--exclude=.snapshot /logs/ /new/logs/
sending incremental file list
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%   95.98MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   49.45MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   44.68MB/s    0:00:01 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   68.03MB/s    0:00:00 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
```

```
sent 527098156 bytes  received 278 bytes  95836078.91 bytes/sec
total size is 527032832  speedup is 1.00
```

5. Démontez les anciens systèmes de fichiers et déplacez les données copiées. Ceci peut être effectué plus rapidement en utilisant un script, comme indiqué dans la "Remplacer le système de fichiers".

```
[root@host1 current]# ./swap.fs.pl /logs,/new/logs
/new/logs unmounted
/logs unmounted
Updated /logs mounted
[root@host1 current]# ./swap.fs.pl /oradata,/new/oradata
/new/oradata unmounted
/oradata unmounted
Updated /oradata mounted
```

6. Vérifiez que les nouveaux systèmes de fichiers sont en place.

```
ontap-nfs1:/host1_logs_new on /logs type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
ontap-nfs1:/host1_oradata_new on /oradata type nfs
(rw,bg,vers=3,rsize=65536,wsize=65536,addr=172.20.101.10)
```

7. Démarrez la base de données.

```
[root@host1 current]# ./dbstart.pl NTAP
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 16:10:07 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
```

Mise en service entièrement automatisée

Cet exemple de script accepte les arguments du SID de la base de données suivis de paires de systèmes de fichiers délimitées par des points communs. Pour l'exemple ci-dessus, la commande est émise comme suit :

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
```

Lorsqu'il est exécuté, l'exemple de script tente d'exécuter la séquence suivante. Il se termine s'il rencontre une erreur dans une étape :

1. Arrêtez la base de données.
2. Convertissez les systèmes de fichiers actuels en mode lecture seule.
3. Utilisez chaque paire d'arguments de système de fichiers délimités par des virgules et synchronisez le premier système de fichiers avec le second.
4. Démonter les systèmes de fichiers précédents.
5. Mettez à jour le `/etc/fstab` classer comme suit :
   a. Créez une sauvegarde à `/etc/fstab.bak`.
   b. Commenter les entrées précédentes pour les systèmes de fichiers antérieurs et nouveaux.
   c. Créez une nouvelle entrée pour le nouveau système de fichiers qui utilise l'ancien point de montage.
6. Montez les systèmes de fichiers.
7. Démarrez la base de données.

Le texte suivant fournit un exemple d'exécution pour ce script :

```
[root@host1 current]# ./migrate.oracle.fs.pl NTAP /logs,/new/logs
/oradata,/new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:05:50 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
SQL> Database closed.
Database dismounted.
ORACLE instance shut down.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP shut down
sending incremental file list
```

```
./
NTAP/
NTAP/1_22_897068759.dbf
    45523968 100%  185.40MB/s    0:00:00 (xfer#1, to-check=15/18)
NTAP/1_23_897068759.dbf
    40601088 100%   81.34MB/s    0:00:00 (xfer#2, to-check=14/18)
...
NTAP/redo/redo02.log
    52429312 100%   70.42MB/s    0:00:00 (xfer#12, to-check=1/18)
NTAP/redo/redo03.log
    52429312 100%   47.08MB/s    0:00:01 (xfer#13, to-check=0/18)
Number of files: 18
Number of files transferred: 13
Total file size: 527032832 bytes
Total transferred file size: 527032832 bytes
Literal data: 527032832 bytes
Matched data: 0 bytes
File list size: 413
File list generation time: 0.001 seconds
File list transfer time: 0.000 seconds
Total bytes sent: 527098156
Total bytes received: 278
sent 527098156 bytes  received 278 bytes  150599552.57 bytes/sec
total size is 527032832  speedup is 1.00
Succesfully replicated filesystem /logs to /new/logs
sending incremental file list
./
NTAP/
NTAP/IOPS.dbf
 10737426432 100%  176.55MB/s    0:00:58 (xfer#1, to-check=10/13)
NTAP/iops.dbf.zip
    22823573 100%    9.48MB/s    0:00:02 (xfer#2, to-check=9/13)
... NTAP/undotbs01.dbf
   309338112 100%   70.76MB/s    0:00:04 (xfer#9, to-check=2/13)
NTAP/undotbs02.dbf
  1073750016 100%  187.65MB/s    0:00:05 (xfer#10, to-check=1/13)
NTAP/users01.dbf
     5251072 100%    5.09MB/s    0:00:00 (xfer#11, to-check=0/13)
Number of files: 13
Number of files transferred: 11
Total file size: 18570092218 bytes
Total transferred file size: 18570092218 bytes
Literal data: 18570092218 bytes
Matched data: 0 bytes
File list size: 277
File list generation time: 0.001 seconds
```

```
File list transfer time: 0.000 seconds
Total bytes sent: 18572359828
Total bytes received: 228
sent 18572359828 bytes  received 228 bytes  177725933.55 bytes/sec
total size is 18570092218  speedup is 1.00
Succesfully replicated filesystem /oradata to /new/oradata
swap 0 /logs /new/logs
/new/logs unmounted
/logs unmounted
Mounted updated /logs
Swapped filesystem /logs for /new/logs
swap 1 /oradata /new/oradata
/new/oradata unmounted
/oradata unmounted
Mounted updated /oradata
Swapped filesystem /oradata for /new/oradata
ORACLE_SID = [oracle] ? The Oracle base has been set to /orabin
SQL*Plus: Release 12.1.0.2.0 Production on Thu Dec 3 17:08:59 2015
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> ORACLE instance started.
Total System Global Area  805306368 bytes
Fixed Size                  2929552 bytes
Variable Size             390073456 bytes
Database Buffers          406847488 bytes
Redo Buffers                5455872 bytes
Database mounted.
Database opened.
SQL> Disconnected from Oracle Database 12c Enterprise Edition Release
12.1.0.2.0 - 64bit Production
With the Partitioning, OLAP, Advanced Analytics and Real Application
Testing options
NTAP started
[root@host1 current]#
```

Migration Oracle ASM spfile et passwd

Le fichier spfile spécifique à ASM et le fichier de mots de passe constituent une difficulté pour terminer la migration impliquant ASM. Par défaut, ces fichiers de métadonnées critiques sont créés sur le premier groupe de disques ASM défini. Si un groupe de disques ASM particulier doit être évacué et supprimé, le fichier spfile et le fichier de mot de passe qui régissent cette instance ASM doivent être déplacés.

Un autre cas d'utilisation où il peut être nécessaire de déplacer ces fichiers est le cas lors du déploiement d'un logiciel de gestion de base de données, tel que SnapManager pour Oracle ou le plug-in SnapCenter pour Oracle. L'une des fonctionnalités de ces produits consiste à restaurer rapidement une base de données en rétablissant l'état des LUN ASM qui hébergent les fichiers de données. Pour ce faire, vous devez mettre le groupe de disques ASM hors ligne avant d'effectuer une restauration. Ce n'est pas un problème tant que les fichiers de données d'une base de données donnée sont isolés dans un groupe de disques ASM dédié.

Lorsque ce groupe de disques contient également le fichier ASM spfile/passwd, la seule façon de mettre le groupe de disques hors ligne est d'arrêter l'instance ASM entière. Il s'agit d'un processus perturbateur, ce qui signifie que le fichier spfile/passwd doit être déplacé.

De production

1. SID de base de données = TOAST
2. Fichiers de données actuels sur `+DATA`
3. Fichiers journaux et fichiers de contrôle actuels sur `+LOGS`
4. Nouveaux groupes de disques ASM définis en tant que `+NEWDATA` et `+NEWLOGS`

Emplacements des fichiers spfile/passwd ASM

La migration de ces fichiers peut s'effectuer sans interruption. Cependant, pour des raisons de sécurité, NetApp recommande de fermer l'environnement de base de données afin de vous assurer que les fichiers ont été déplacés et que la configuration est correctement mise à jour. Cette procédure doit être répétée si plusieurs instances ASM sont présentes sur un serveur.

Identifier les instances ASM

Identifier les instances ASM en fonction des données enregistrées dans le `oratab` fichier. Les instances ASM sont signalées par un symbole +.

```
-bash-4.1$ cat /etc/oratab | grep '^+'
+ASM:/orabin/grid:N             # line added by Agent
```

Il existe une instance ASM appelée +ASM sur ce serveur.

Assurez-vous que toutes les bases de données sont arrêtées

Le seul processus smon visible doit être le smon de l'instance ASM utilisée. La présence d'un autre processus smon indique qu'une base de données est toujours en cours d'exécution.

```
-bash-4.1$ ps -ef | grep smon
oracle     857     1  0 18:26 ?        00:00:00 asm_smon_+ASM
```

Le seul processus smon est l'instance ASM elle-même. Cela signifie qu'aucune autre base de données n'est en cours d'exécution et que vous pouvez continuer en toute sécurité sans risque d'interruption des opérations de la base de données.

Localisez les fichiers

Identifiez l'emplacement actuel du fichier spfile et du fichier de mots de passe ASM à l'aide du `spget` et `pwget` commandes.

```
bash-4.1$ asmcmd
ASMCMD> spget
+DATA/spfile.ora
```

```
ASMCMD> pwget --asm
+DATA/orapwasm
```

Les fichiers se trouvent tous deux à la base du `+DATA` groupe de disques.

Copier des fichiers

Copiez les fichiers dans le nouveau groupe de disques ASM avec le `spcopy` et `pwcopy` commandes. Si le nouveau groupe de disques a été créé récemment et est actuellement vide, il peut être nécessaire de le monter en premier.

```
ASMCMD> mount NEWDATA
```

```
ASMCMD> spcopy +DATA/spfile.ora +NEWDATA/spfile.ora
copying +DATA/spfile.ora -> +NEWDATA/spfilea.ora
```

```
ASMCMD> pwcopy +DATA/orapwasm +NEWDATA/orapwasm
copying +DATA/orapwasm -> +NEWDATA/orapwasm
```

Les fichiers ont été copiés depuis `+DATA` à `+NEWDATA`.

Mettre à jour l'instance ASM

L'instance ASM doit maintenant être mise à jour pour refléter le changement d'emplacement. Le `spset` et `pwset` Les commandes mettent à jour les métadonnées ASM requises pour démarrer le groupe de disques ASM.

```
ASMCMD> spset +NEWDATA/spfile.ora
ASMCMD> pwset --asm +NEWDATA/orapwasm
```

Activez ASM à l'aide de fichiers mis à jour

À ce stade, l'instance ASM utilise toujours les emplacements précédents de ces fichiers. L'instance doit être redémarrée pour forcer une relecture des fichiers à partir de leurs nouveaux emplacements et pour libérer les verrous sur les fichiers précédents.

```
-bash-4.1$ sqlplus / as sysasm
SQL> shutdown immediate;
ASM diskgroups volume disabled
ASM diskgroups dismounted
ASM instance shutdown
```

```
SQL> startup
ASM instance started
Total System Global Area 1140850688 bytes
Fixed Size                  2933400 bytes
Variable Size            1112751464 bytes
ASM Cache                  25165824 bytes
ORA-15032: not all alterations performed
ORA-15017: diskgroup "NEWDATA" cannot be mounted
ORA-15013: diskgroup "NEWDATA" is already mounted
```

Supprimez les anciens fichiers spfile et les anciens fichiers de mots de passe

Si la procédure a été effectuée avec succès, les fichiers précédents ne sont plus verrouillés et peuvent maintenant être supprimés.

```
-bash-4.1$ asmcmd
ASMCMD> rm +DATA/spfile.ora
ASMCMD> rm +DATA/orapwasm
```

Copie d'Oracle ASM vers ASM

Oracle ASM est essentiellement un gestionnaire de volumes combiné léger et un système de fichiers. Comme le système de fichiers n'est pas facilement visible, RMAN doit être utilisé pour effectuer des opérations de copie. Même si un processus de migration basé sur la copie est sûr et simple, il provoque certaines perturbations. Les interruptions peuvent être minimisées, mais pas totalement éliminées.

Si vous souhaitez effectuer une migration sans interruption d'une base de données ASM, il est préférable d'exploiter la capacité d'ASM à rééquilibrer les extensions ASM vers de nouveaux LUN lors de la suppression des anciennes LUN. Cette opération est généralement sûre et non disruptive, mais elle n'offre pas de chemin « back-out ». En cas de problèmes fonctionnels ou de performances, la seule option consiste à migrer les données vers la source.

Ce risque peut être évité en copiant la base de données vers le nouvel emplacement plutôt que de déplacer les données, afin que les données d'origine ne soient pas modifiées. La base de données peut être entièrement testée à son nouvel emplacement avant la mise en service, et la base de données d'origine est disponible comme option de retour en arrière si des problèmes sont détectés.

Cette procédure est l'une des nombreuses options impliquant RMAN. Il est conçu pour permettre un processus en deux étapes dans lequel la sauvegarde initiale est créée, puis synchronisée par la suite via la relecture du journal. Ce processus est recommandé pour réduire les temps d'indisponibilité, car il permet à la base de données de rester opérationnelle et d'assurer l'accès aux données pendant la copie de base initiale.

Copier la base de données

Oracle RMAN crée une copie de niveau 0 (complète) de la base de données source actuellement située sur le groupe de disques ASM `+DATA` vers le nouvel emplacement sur `+NEWDATA`.

```
-bash-4.1$ rman target /
Recovery Manager: Release 12.1.0.2.0 - Production on Sun Dec 6 17:40:03
2015
Copyright (c) 1982, 2014, Oracle and/or its affiliates.  All rights
reserved.
connected to target database: TOAST (DBID=2084313411)
RMAN> backup as copy incremental level 0 database format '+NEWDATA' tag
'ONTAP_MIGRATION';
Starting backup at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=302 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
...
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
output file name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
tag=ONTAP_MIGRATION RECID=5 STAMP=897759622
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWDATA/TOAST/BACKUPSET/2015_12_06/nnsnn0_ontap_migration_0.262.89
7759623 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

Forcer le changement de journal d’archivage

Vous devez forcer un commutateur de journal d’archivage pour vous assurer que les journaux d’archivage contiennent toutes les données nécessaires pour que la copie soit totalement cohérente. Sans cette commande, les données clés peuvent toujours être présentes dans les journaux de reprise.

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

Arrêtez la base de données source

L'interruption commence à cette étape car la base de données est arrêtée et placée en mode lecture seule à accès limité. Pour arrêter la base de données source, exécutez les commandes suivantes :

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                390073456 bytes
Database Buffers             406847488 bytes
Redo Buffers                   5455872 bytes
```

Sauvegarde Controlfile

Vous devez sauvegarder le fichier de contrôle si vous devez abandonner la migration et revenir à l'emplacement de stockage d'origine. Une copie du fichier de contrôle de sauvegarde n'est pas nécessaire à 100 %, mais elle facilite le processus de réinitialisation des emplacements des fichiers de base de données vers leur emplacement d'origine.

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=358 device type=DISK
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151206T174753 RECID=6
STAMP=897760073
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

Mises à jour des paramètres

Le fichier spfile actuel contient des références aux fichiers de contrôle sur leurs emplacements actuels dans l'ancien groupe de disques ASM. Il doit être édité, ce qui est facile à faire en éditant une version intermédiaire de pfile.

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

Mettre à jour le fichier pfile

Mettez à jour tous les paramètres faisant référence aux anciens groupes de disques ASM pour refléter les nouveaux noms de groupes de disques ASM. Enregistrez ensuite le fichier pfile mis à jour. Assurez-vous que le `db_create` des paramètres sont présents.

Dans l'exemple ci-dessous, les références à `+DATA` ils ont été remplacés par `+NEWDATA` sont surlignés en jaune. Deux paramètres clés sont le `db_create` paramètres qui créent de nouveaux fichiers à l'emplacement correct.

```
*.compatible='12.1.0.2.0'
*.control_files='+NEWLOGS/TOAST/CONTROLFILE/current.258.897683139'
*.db_block_size=8192
*. db_create_file_dest='+NEWDATA'
*. db_create_online_log_dest_1='+NEWLOGS'
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=+NEWLOGS'
*.log_archive_format='%t_%s_%r.dbf'
```

Mettre à jour le fichier init.ora

La plupart des bases de données ASM utilisent un `init.ora` fichier situé dans le `$ORACLE_HOME/dbs` Répertoire, qui est un point vers le fichier spfile sur le groupe de disques ASM. Ce fichier doit être redirigé vers un emplacement du nouveau groupe de disques ASM.

```
-bash-4.1$ cd $ORACLE_HOME/dbs
-bash-4.1$ cat initTOAST.ora
SPFILE='+DATA/TOAST/spfileTOAST.ora'
```

Modifiez ce fichier comme suit :

```
SPFILE=+NEWLOGS/TOAST/spfileTOAST.ora
```

Récréation du fichier de paramètres

Le fichier spfile est maintenant prêt à être rempli par les données du fichier pfile modifié.

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

Démarrez la base de données pour commencer à utiliser le nouveau fichier spfile

Démarrez la base de données pour vous assurer qu'elle utilise maintenant le fichier spfile nouvellement créé et que toute autre modification des paramètres système est correctement enregistrée.

```
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                373296240 bytes
Database Buffers             423624704 bytes
Redo Buffers                   5455872 bytes
```

Restaurer le fichier de contrôle

Le fichier de contrôle de sauvegarde créé par RMAN peut également être restauré directement par RMAN à l'emplacement spécifié dans le nouveau fichier spfile.

```
RMAN> restore controlfile from
'+DATA/TOAST/CONTROLFILE/current.258.897683139';
Starting restore at 06-DEC-15
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=417 device type=DISK
channel ORA_DISK_1: copied control file copy
output file name=+NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
Finished restore at 06-DEC-15
```

Montez la base de données et vérifiez l'utilisation du nouveau fichier de contrôle.

```
RMAN> alter database mount;
using target database control file instead of recovery catalog
Statement processed
```

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
+NEWLOGS/TOAST/CONTROLFILE/cur
                                                 rent.273.897761061
```

Relecture du journal

La base de données utilise actuellement les fichiers de données dans l'ancien emplacement. Avant de pouvoir utiliser la copie, elles doivent être synchronisées. Le temps s'est écoulé pendant le processus de copie initial et les modifications ont été enregistrées principalement dans les journaux d'archivage. Ces modifications sont répliquées comme suit :

1. Effectuez une sauvegarde incrémentielle RMAN contenant les journaux d'archivage.

```
RMAN> backup incremental level 1 format '+NEWLOGS' for recover of copy
with tag 'ONTAP_MIGRATION' database;
Starting backup at 06-DEC-15
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=62 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001
name=+DATA/TOAST/DATAFILE/system.262.897683141
input datafile file number=00002
name=+DATA/TOAST/DATAFILE/sysaux.260.897683143
input datafile file number=00003
name=+DATA/TOAST/DATAFILE/undotbs1.257.897683145
input datafile file number=00004
name=+DATA/TOAST/DATAFILE/users.264.897683151
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current control file in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 06-DEC-15
channel ORA_DISK_1: finished piece 1 at 06-DEC-15
piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/ncsnn1_ontap_migration_0.267.
897762697 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 06-DEC-15
```

2. Relire le journal.

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 06-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001
name=+NEWDATA/TOAST/DATAFILE/system.259.897759609
recovering datafile copy file number=00002
name=+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
recovering datafile copy file number=00003
name=+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
recovering datafile copy file number=00004
name=+NEWDATA/TOAST/DATAFILE/users.258.897759623
channel ORA_DISK_1: reading from backup piece
+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.8977626
93
channel ORA_DISK_1: piece
handle=+NEWLOGS/TOAST/BACKUPSET/2015_12_06/nnndn1_ontap_migration_0.268.
897762693 tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

Activation

Le fichier de contrôle restauré fait toujours référence aux fichiers de données à l'emplacement d'origine et contient également les informations de chemin des fichiers de données copiés.

1. Pour modifier les fichiers de données actifs, exécutez `switch database to copy` commande.

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/system.259.897759609"
datafile 2 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615"
datafile 3 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619"
datafile 4 switched to datafile copy
"+NEWDATA/TOAST/DATAFILE/users.258.897759623"
```

   Les fichiers de données actifs sont désormais les fichiers de données copiés, mais des modifications peuvent encore être contenues dans les journaux de reprise finaux.

2. Pour relire tous les journaux restants, exécutez le `recover database` commande. Si le message s'affiche `media recovery complete` apparaît, le processus a réussi.

```
RMAN> recover database;
Starting recover at 06-DEC-15
using channel ORA_DISK_1
starting media recovery
media recovery complete, elapsed time: 00:00:01
Finished recover at 06-DEC-15
```

Ce processus a uniquement modifié l'emplacement des fichiers de données normaux. Les fichiers de données temporaires doivent être renommés, mais ils n'ont pas besoin d'être copiés car ils sont temporaires uniquement. La base de données est actuellement inactive, il n'y a donc pas de données actives dans les fichiers de données temporaires.

3. Pour déplacer les fichiers de données temporaires, identifiez d'abord leur emplacement.

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
------------------------------------------------------------------------
--------
1 +DATA/TOAST/TEMPFILE/temp.263.897683145
```

4. Déplacez les fichiers de données temporaires à l'aide d'une commande RMAN qui définit le nouveau nom de chaque fichier de données. Avec Oracle Managed Files (OMF), le nom complet n'est pas nécessaire ; le groupe de disques ASM est suffisant. Lorsque la base de données est ouverte, OMF est lié à l'emplacement approprié sur le groupe de disques ASM. Pour déplacer des fichiers, exécutez les commandes suivantes :

```
run {
set newname for tempfile 1 to '+NEWDATA';
switch tempfile all;
}
```

```
RMAN> run {
2> set newname for tempfile 1 to '+NEWDATA';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to +NEWDATA in control file
```

Migration du journal de reprise

Le processus de migration est presque terminé, mais les journaux de reprise se trouvent toujours sur le groupe de disques ASM d'origine. Les journaux de reprise ne peuvent pas être transférés directement. Un nouvel ensemble de journaux de reprise est créé et ajouté à la configuration, suivi d'un DROP des anciens journaux.

1. Identifiez le nombre de groupes de fichiers redo log et leurs numéros de groupe respectifs.

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
---------------------------------------------------------------------------
--------
1 +DATA/TOAST/ONLINELOG/group_1.261.897683139
2 +DATA/TOAST/ONLINELOG/group_2.259.897683139
3 +DATA/TOAST/ONLINELOG/group_3.256.897683139
```

2. Indiquez la taille des journaux de reprise.

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
---------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Pour chaque journal de reprise, créez un groupe avec une configuration correspondante. Si vous n'utilisez pas OMF, vous devez spécifier le chemin complet. C'est également un exemple qui utilise le `db_create_online_log` paramètres. Comme indiqué précédemment, ce paramètre a été défini sur +NEWLOGS. Cette configuration vous permet d'utiliser les commandes suivantes pour créer de nouveaux journaux en ligne sans avoir à spécifier un emplacement de fichier ou même un groupe de disques ASM spécifique.

```
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
RMAN> alter database add logfile size 52428800;
Statement processed
```

4. Ouvrez la base de données.

```
SQL> alter database open;
Database altered.
```

5. Supprimez les anciens journaux.

```
RMAN> alter database drop logfile group 1;
Statement processed
```

6. Si vous rencontrez une erreur qui vous empêche de supprimer un journal actif, forcez un commutateur au journal suivant pour libérer le verrouillage et forcer un point de contrôle global. Un exemple est illustré ci-dessous. La tentative de suppression du groupe de fichiers journaux 3, qui se trouvait sur l'ancien emplacement, a été refusée parce qu'il y avait encore des données actives dans ce fichier journal. Un archivage de journaux après un point de contrôle vous permet de supprimer le fichier journal.

```
RMAN> alter database drop logfile group 3;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 3 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 3 thread 1:
'+LOGS/TOAST/ONLINELOG/group_3.259.897563549'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 3;
Statement processed
```

7. Vérifiez l'environnement pour vous assurer que tous les paramètres basés sur l'emplacement sont mis à jour.

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

8. Le script suivant explique comment simplifier ce processus :

```
[root@host1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
+NEWDATA/TOAST/DATAFILE/system.259.897759609
+NEWDATA/TOAST/DATAFILE/sysaux.263.897759615
+NEWDATA/TOAST/DATAFILE/undotbs1.264.897759619
+NEWDATA/TOAST/DATAFILE/users.258.897759623
TOAST redo logs:
+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123
+NEWLOGS/TOAST/ONLINELOG/group_5.265.897763125
+NEWLOGS/TOAST/ONLINELOG/group_6.264.897763125
TOAST temp datafiles:
+NEWDATA/TOAST/TEMPFILE/temp.260.897763165
TOAST spfile
spfile                                   string
+NEWDATA/spfiletoast.ora
TOAST key parameters
control_files +NEWLOGS/TOAST/CONTROLFILE/current.273.897761061
log_archive_dest_1 LOCATION=+NEWLOGS
db_create_file_dest +NEWDATA
db_create_online_log_dest_1 +NEWLOGS
```

9. Si les groupes de disques ASM ont été complètement évacués, ils peuvent maintenant être démontés avec `asmcmd`. Cependant, dans de nombreux cas, les fichiers appartenant à d'autres bases de données ou au fichier ASM spfile/passwd peuvent toujours être présents.

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

Copie d'Oracle ASM vers le système de fichiers

La procédure de copie d'Oracle ASM vers un système de fichiers est très similaire à la procédure de copie d'ASM vers ASM, avec des avantages et des restrictions similaires. La différence principale est la syntaxe des différentes commandes et paramètres de configuration lors de l'utilisation d'un système de fichiers visible par opposition à un groupe de disques ASM.

Copier la base de données

Oracle RMAN permet de créer une copie de niveau 0 (complète) de la base de données source actuellement située sur le groupe de disques ASM `+DATA` vers le nouvel emplacement sur `/oradata`.

```
RMAN> backup as copy incremental level 0 database format
'/oradata/TOAST/%U' tag 'ONTAP_MIGRATION';
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=377 device type=DISK
channel ORA_DISK_1: starting datafile copy
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-
1_01r5fhjg tag=ONTAP_MIGRATION RECID=1 STAMP=911722099
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-
2_02r5fhjo tag=ONTAP_MIGRATION RECID=2 STAMP=911722106
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt tag=ONTAP_MIGRATION RECID=3 STAMP=911722113
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:07
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/oradata/TOAST/cf_D-TOAST_id-2098173325_04r5fhk5
tag=ONTAP_MIGRATION RECID=4 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting datafile copy
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
output file name=/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-
4_05r5fhk6 tag=ONTAP_MIGRATION RECID=5 STAMP=911722118
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
channel ORA_DISK_1: starting incremental level 0 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
including current SPFILE in backup set
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/oradata/TOAST/06r5fhk7_1_1 tag=ONTAP_MIGRATION comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
```

Forcer le changement de journal d’archivage

Forcer le commutateur de journal d’archivage est nécessaire pour s’assurer que les journaux d’archivage contiennent toutes les données requises pour rendre la copie entièrement cohérente. Sans cette commande, les données clés peuvent toujours être présentes dans les journaux de reprise. Pour forcer un commutateur de journal d’archivage, exécutez la commande suivante :

```
RMAN> sql 'alter system archive log current';
sql statement: alter system archive log current
```

Arrêtez la base de données source

L'interruption commence à cette étape car la base de données est arrêtée et placée en mode lecture seule à accès limité. Pour arrêter la base de données source, exécutez les commandes suivantes :

```
RMAN> shutdown immediate;
using target database control file instead of recovery catalog
database closed
database dismounted
Oracle instance shut down
RMAN> startup mount;
connected to target database (not started)
Oracle instance started
database mounted
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

Sauvegarde Controlfile

Sauvegarder les fichiers de contrôle si vous devez abandonner la migration et revenir à l'emplacement de stockage d'origine. Une copie du fichier de contrôle de sauvegarde n'est pas nécessaire à 100 %, mais elle facilite le processus de réinitialisation des emplacements des fichiers de base de données vers leur emplacement d'origine.

```
RMAN> backup as copy current controlfile format '/tmp/TOAST.ctrl';
Starting backup at 08-DEC-15
using channel ORA_DISK_1
channel ORA_DISK_1: starting datafile copy
copying current control file
output file name=/tmp/TOAST.ctrl tag=TAG20151208T194540 RECID=30
STAMP=897939940
channel ORA_DISK_1: datafile copy complete, elapsed time: 00:00:01
Finished backup at 08-DEC-15
```

Mises à jour des paramètres

```
RMAN> create pfile='/tmp/pfile' from spfile;
Statement processed
```

Mettre à jour le fichier pfile

Tous les paramètres faisant référence aux anciens groupes de disques ASM doivent être mis à jour et, dans certains cas, supprimés lorsqu'ils ne sont plus pertinents. Mettez-les à jour pour refléter les nouveaux chemins du système de fichiers et enregistrez le fichier pfile mis à jour. Assurez-vous que le chemin cible complet est répertorié. Pour mettre à jour ces paramètres, exécutez les commandes suivantes :

```
*.audit_file_dest='/orabin/admin/TOAST/adump'
*.audit_trail='db'
*.compatible='12.1.0.2.0'
*.control_files='/logs/TOAST/arch/control01.ctl','/logs/TOAST/redo/control
02.ctl'
*.db_block_size=8192
*.db_domain=''
*.db_name='TOAST'
*.diagnostic_dest='/orabin'
*.dispatchers='(PROTOCOL=TCP) (SERVICE=TOASTXDB)'
*.log_archive_dest_1='LOCATION=/logs/TOAST/arch'
*.log_archive_format='%t_%s_%r.dbf'
*.open_cursors=300
*.pga_aggregate_target=256m
*.processes=300
*.remote_login_passwordfile='EXCLUSIVE'
*.sga_target=768m
*.undo_tablespace='UNDOTBS1'
```

Désactivez le fichier init.ora d'origine

Ce fichier se trouve dans le `$ORACLE_HOME/dbs` Et se trouve généralement dans un fichier pfile qui sert de pointeur vers le fichier spfile sur le groupe de disques ASM. Pour vous assurer que le fichier spfile d'origine n'est plus utilisé, renommez-le. Ne le supprimez pas, cependant, car ce fichier est nécessaire si la migration doit être abandonnée.

```
[oracle@jfsc1 ~]$ cd $ORACLE_HOME/dbs
[oracle@jfsc1 dbs]$ cat initTOAST.ora
SPFILE='+ASM0/TOAST/spfileTOAST.ora'
[oracle@jfsc1 dbs]$ mv initTOAST.ora initTOAST.ora.prev
[oracle@jfsc1 dbs]$
```

Récréation du fichier de paramètres

Il s'agit de la dernière étape de la relocalisation de fichier spfile. Le fichier spfile d'origine n'est plus utilisé et la base de données est actuellement démarrée (mais pas montée) à l'aide du fichier intermédiaire. Le contenu de ce fichier peut être écrit dans le nouvel emplacement spfile comme suit :

```
RMAN> create spfile from pfile='/tmp/pfile';
Statement processed
```

Démarrez la base de données pour commencer à utiliser le nouveau fichier spfile

Vous devez démarrer la base de données pour libérer les verrous sur le fichier intermédiaire et démarrer la base de données en utilisant uniquement le nouveau fichier spfile. Le démarrage de la base de données prouve également que le nouvel emplacement spfile est correct et que ses données sont valides.

```
RMAN> shutdown immediate;
Oracle instance shut down
RMAN> startup nomount;
connected to target database (not started)
Oracle instance started
Total System Global Area     805306368 bytes
Fixed Size                     2929552 bytes
Variable Size                331353200 bytes
Database Buffers             465567744 bytes
Redo Buffers                   5455872 bytes
```

Restaurer le fichier de contrôle

Un fichier de contrôle de sauvegarde a été créé au niveau du chemin `/tmp/TOAST.ctrl` plus tôt dans la procédure. Le nouveau fichier spfile définit les emplacements des fichiers de contrôle comme `/logfs/TOAST/ctrl/ctrlfile1.ctrl` et `/logfs/TOAST/redo/ctrlfile2.ctrl`. Cependant, ces fichiers n'existent pas encore.

1. Cette commande restaure les données du fichier de contrôle dans les chemins définis dans le fichier spfile.

   ```
   RMAN> restore controlfile from '/tmp/TOAST.ctrl';
   Starting restore at 13-MAY-16
   using channel ORA_DISK_1
   channel ORA_DISK_1: copied control file copy
   output file name=/logs/TOAST/arch/control01.ctl
   output file name=/logs/TOAST/redo/control02.ctl
   Finished restore at 13-MAY-16
   ```

2. Exécutez la commande mount pour que les fichiers de contrôle soient correctement découverts et contiennent des données valides.

```
RMAN> alter database mount;
Statement processed
released channel: ORA_DISK_1
```

Pour valider le `control_files` paramètre, exécutez la commande suivante :

```
SQL> show parameter control_files;
NAME                                 TYPE        VALUE
------------------------------------ -----------
------------------------------
control_files                        string
/logs/TOAST/arch/control01.ctl
                                                 ,
/logs/TOAST/redo/control02.c
                                                 tl
```

Relecture du journal

La base de données utilise actuellement les fichiers de données dans l'ancien emplacement. Avant de pouvoir utiliser la copie, les fichiers de données doivent être synchronisés. Le temps s'est écoulé pendant le processus de copie initial et les modifications ont été enregistrées principalement dans les journaux d'archivage. Ces modifications sont répliquées dans les deux étapes suivantes.

1. Effectuez une sauvegarde incrémentielle RMAN contenant les journaux d'archivage.

```
RMAN>  backup incremental level 1 format '/logs/TOAST/arch/%U' for
recover of copy with tag 'ONTAP_MIGRATION' database;
Starting backup at 13-MAY-16
using target database control file instead of recovery catalog
allocated channel: ORA_DISK_1
channel ORA_DISK_1: SID=124 device type=DISK
channel ORA_DISK_1: starting incremental level 1 datafile backup set
channel ORA_DISK_1: specifying datafile(s) in backup set
input datafile file number=00001 name=+ASM0/TOAST/system01.dbf
input datafile file number=00002 name=+ASM0/TOAST/sysaux01.dbf
input datafile file number=00003 name=+ASM0/TOAST/undotbs101.dbf
input datafile file number=00004 name=+ASM0/TOAST/users01.dbf
channel ORA_DISK_1: starting piece 1 at 13-MAY-16
channel ORA_DISK_1: finished piece 1 at 13-MAY-16
piece handle=/logs/TOAST/arch/09r5fj8i_1_1 tag=ONTAP_MIGRATION
comment=NONE
channel ORA_DISK_1: backup set complete, elapsed time: 00:00:01
Finished backup at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

2. Relire les journaux.

```
RMAN> recover copy of database with tag 'ONTAP_MIGRATION';
Starting recover at 13-MAY-16
using channel ORA_DISK_1
channel ORA_DISK_1: starting incremental datafile backup set restore
channel ORA_DISK_1: specifying datafile copies to recover
recovering datafile copy file number=00001 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
recovering datafile copy file number=00002 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
recovering datafile copy file number=00003 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
recovering datafile copy file number=00004 name=/oradata/TOAST/data_D-
TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
channel ORA_DISK_1: reading from backup piece
/logs/TOAST/arch/09r5fj8i_1_1
channel ORA_DISK_1: piece handle=/logs/TOAST/arch/09r5fj8i_1_1
tag=ONTAP_MIGRATION
channel ORA_DISK_1: restored backup piece 1
channel ORA_DISK_1: restore complete, elapsed time: 00:00:01
Finished recover at 13-MAY-16
RMAN-06497: WARNING: control file is not current, control file
AUTOBACKUP skipped
```

Activation

Le fichier de contrôle restauré fait toujours référence aux fichiers de données à l'emplacement d'origine et contient également les informations de chemin des fichiers de données copiés.

1. Pour modifier les fichiers de données actifs, exécutez `switch database to copy` commande :

```
RMAN> switch database to copy;
datafile 1 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSTEM_FNO-1_01r5fhjg"
datafile 2 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-SYSAUX_FNO-2_02r5fhjo"
datafile 3 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt"
datafile 4 switched to datafile copy "/oradata/TOAST/data_D-TOAST_I-
2098173325_TS-USERS_FNO-4_05r5fhk6"
```

2. Bien que les fichiers de données soient parfaitement cohérents, une dernière étape est nécessaire pour relire les modifications restantes enregistrées dans les journaux de reprise en ligne. Utilisez le `recover database` pour relire ces modifications et rendre la copie identique à 100 % à l'original. Toutefois, la copie n'est pas encore ouverte.

```
RMAN> recover database;
Starting recover at 13-MAY-16
using channel ORA_DISK_1
starting media recovery
archived log for thread 1 with sequence 28 is already on disk as file
+ASM0/TOAST/redo01.log
archived log file name=+ASM0/TOAST/redo01.log thread=1 sequence=28
media recovery complete, elapsed time: 00:00:00
Finished recover at 13-MAY-16
```

Déplacer les fichiers de données temporaires

1. Identifiez l'emplacement des fichiers de données temporaires toujours en cours d'utilisation sur le groupe de disques d'origine.

```
RMAN> select file#||' '||name from v$tempfile;
FILE#||''||NAME
------------------------------------------------------------------------
--------
1 +ASM0/TOAST/temp01.dbf
```

2. Pour déplacer les fichiers de données, exécutez les commandes suivantes. S'il existe de nombreux fichiers tempfiles, utilisez un éditeur de texte pour créer la commande RMAN, puis coupez-la et collez-la.

```
RMAN> run {
2> set newname for tempfile 1 to '/oradata/TOAST/temp01.dbf';
3> switch tempfile all;
4> }
executing command: SET NEWNAME
renamed tempfile 1 to /oradata/TOAST/temp01.dbf in control file
```

Migration du journal de reprise

Le processus de migration est presque terminé, mais les journaux de reprise se trouvent toujours sur le groupe de disques ASM d'origine. Les journaux de reprise ne peuvent pas être transférés directement. Un nouvel ensemble de journaux de reprise est créé et ajouté à la configuration, suivant un DROP des anciens journaux.

1. Identifiez le nombre de groupes de fichiers redo log et leurs numéros de groupe respectifs.

```
RMAN> select group#||' '||member from v$logfile;
GROUP#||''||MEMBER
---------------------------------------------------------------------------
--------
1 +ASM0/TOAST/redo01.log
2 +ASM0/TOAST/redo02.log
3 +ASM0/TOAST/redo03.log
```

2. Indiquez la taille des journaux de reprise.

```
RMAN> select group#||' '||bytes from v$log;
GROUP#||''||BYTES
---------------------------------------------------------------------------
--------
1 52428800
2 52428800
3 52428800
```

3. Pour chaque fichier redo log, créez un groupe en utilisant la même taille que le groupe de fichiers redo log actuel à l'aide du nouvel emplacement du système de fichiers.

```
RMAN> alter database add logfile '/logs/TOAST/redo/log00.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log01.rdo' size
52428800;
Statement processed
RMAN> alter database add logfile '/logs/TOAST/redo/log02.rdo' size
52428800;
Statement processed
```

4. Supprimez les anciens groupes de fichiers journaux qui se trouvent toujours sur le stockage précédent.

```
RMAN> alter database drop logfile group 4;
Statement processed
RMAN> alter database drop logfile group 5;
Statement processed
RMAN> alter database drop logfile group 6;
Statement processed
```

5. Si une erreur bloque la suppression d'un journal actif, forcez un commutateur au journal suivant pour libérer le verrouillage et forcer un point de contrôle global. Un exemple est illustré ci-dessous. La tentative de suppression du groupe de fichiers journaux 3, qui se trouvait sur l'ancien emplacement, a été refusée

parce qu'il y avait encore des données actives dans ce fichier journal. L'archivage des journaux suivi d'un point de contrôle permet la suppression des fichiers journaux.

```
RMAN> alter database drop logfile group 4;
RMAN-00571: ===========================================================
RMAN-00569: =============== ERROR MESSAGE STACK FOLLOWS ===============
RMAN-00571: ===========================================================
RMAN-03002: failure of sql statement command at 12/08/2015 20:23:51
ORA-01623: log 4 is current log for instance TOAST (thread 4) - cannot
drop
ORA-00312: online log 4 thread 1:
'+NEWLOGS/TOAST/ONLINELOG/group_4.266.897763123'
RMAN> alter system switch logfile;
Statement processed
RMAN> alter system checkpoint;
Statement processed
RMAN> alter database drop logfile group 4;
Statement processed
```

6. Vérifiez l'environnement pour vous assurer que tous les paramètres basés sur l'emplacement sont mis à jour.

```
SQL> select name from v$datafile;
SQL> select member from v$logfile;
SQL> select name from v$tempfile;
SQL> show parameter spfile;
SQL> select name, value from v$parameter where value is not null;
```

7. Le script suivant explique comment faciliter ce processus.

```
[root@jfsc1 current]# ./checkdbdata.pl TOAST
TOAST datafiles:
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-3_03r5fhjt
/oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
TOAST redo logs:
/logs/TOAST/redo/log00.rdo
/logs/TOAST/redo/log01.rdo
/logs/TOAST/redo/log02.rdo
TOAST temp datafiles:
/oradata/TOAST/temp01.dbf
TOAST spfile
spfile                                 string
/orabin/product/12.1.0/dbhome_
                                                      1/dbs/spfileTOAST.ora
TOAST key parameters
control_files /logs/TOAST/arch/control01.ctl,
/logs/TOAST/redo/control02.ctl
log_archive_dest_1 LOCATION=/logs/TOAST/arch
```

8. Si les groupes de disques ASM ont été complètement évacués, ils peuvent maintenant être démontés avec `asmcmd`. Dans de nombreux cas, les fichiers appartenant à d'autres bases de données ou au fichier ASM spfile/passwd peuvent toujours être présents.

```
-bash-4.1$ . oraenv
ORACLE_SID = [TOAST] ? +ASM
The Oracle base remains unchanged with value /orabin
-bash-4.1$ asmcmd
ASMCMD> umount DATA
ASMCMD>
```

Procédure de nettoyage du fichier de données

Le processus de migration peut donner lieu à des fichiers de données avec une syntaxe longue ou chiffrée, selon la façon dont Oracle RMAN a été utilisé. Dans l'exemple illustré ici, la sauvegarde a été effectuée avec le format de fichier de `/oradata/TOAST/%U`. `%U` Indique que RMAN doit créer un nom unique par défaut pour chaque fichier de données. Le résultat est similaire à ce qui est affiché dans le texte suivant. Les noms traditionnels des fichiers de données sont incorporés dans les noms. Pour ce faire, utilisez l'approche par script illustrée à la "Nettoyage de migration ASM".

```
[root@jfsc1 current]# ./fixuniquenames.pl TOAST
#sqlplus Commands
shutdown immediate;
startup mount;
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSTEM_FNO-1_01r5fhjg
/oradata/TOAST/system.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-SYSAUX_FNO-2_02r5fhjo
/oradata/TOAST/sysaux.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-UNDOTBS1_FNO-
3_03r5fhjt /oradata/TOAST/undotbs1.dbf
host mv /oradata/TOAST/data_D-TOAST_I-2098173325_TS-USERS_FNO-4_05r5fhk6
/oradata/TOAST/users.dbf
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSTEM_FNO-1_01r5fhjg' to '/oradata/TOAST/system.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
SYSAUX_FNO-2_02r5fhjo' to '/oradata/TOAST/sysaux.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
UNDOTBS1_FNO-3_03r5fhjt' to '/oradata/TOAST/undotbs1.dbf';
alter database rename file '/oradata/TOAST/data_D-TOAST_I-2098173325_TS-
USERS_FNO-4_05r5fhk6' to '/oradata/TOAST/users.dbf';
alter database open;
```

Rééquilibrage d'Oracle ASM

Comme nous l'avons vu précédemment, un groupe de disques Oracle ASM peut être migré en toute transparence vers un nouveau système de stockage en utilisant le processus de rééquilibrage. En résumé, le processus de rééquilibrage nécessite l'ajout de LUN de taille égale au groupe existant de LUN, suivi d'une opération de DROP de la LUN précédente. Oracle ASM déplace automatiquement les données sous-jacentes vers un nouveau stockage selon une disposition optimale, puis libère les anciens LUN une fois l'opération terminée.

Le processus de migration utilise des E/S séquentielles efficaces et ne provoque généralement aucune interruption des performances. En revanche, le taux de migration peut être ralenti lorsque cela est nécessaire.

Identifiez les données à migrer

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
SQL> select group_number||' '||name from v$asm_diskgroup;
1 NEWDATA
```

Créer des LUN

Créez de nouvelles LUN de la même taille et définissez l'appartenance des utilisateurs et des groupes selon les besoins. Les LUN doivent s'afficher comme `CANDIDATE` disques.

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
 0 0 /dev/mapper/3600a098038303537762b47594c31586b CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315869 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c315858 CANDIDATE
 0 0 /dev/mapper/3600a098038303537762b47594c31586a CANDIDATE
NEWDATA_0003 1 10240 /dev/mapper/3600a098038303537762b47594c315864 MEMBER
NEWDATA_0002 1 10240 /dev/mapper/3600a098038303537762b47594c315863 MEMBER
NEWDATA_0000 1 10240 /dev/mapper/3600a098038303537762b47594c315861 MEMBER
NEWDATA_0001 1 10240 /dev/mapper/3600a098038303537762b47594c315862 MEMBER
```

Ajouter de nouvelles LUN

Même si les opérations d'ajout et de suppression peuvent être effectuées ensemble, il est généralement plus facile d'ajouter de nouvelles LUN en deux étapes. Commencez par ajouter les nouvelles LUN au groupe de disques. Cette étape entraîne la migration de la moitié des extensions des LUN ASM actuelles vers les nouvelles LUN.

La puissance de rééquilibrage indique la vitesse à laquelle les données sont transférées. Plus le nombre est élevé, plus le parallélisme du transfert de données est élevé. La migration s'effectue au moyen d'opérations d'E/S séquentielles efficaces, peu susceptibles d'entraîner des problèmes de performances. Toutefois, si nécessaire, le pouvoir de rééquilibrage d'une migration en cours peut être ajusté avec le `alter diskgroup [name] rebalance power [level]` commande. Les migrations types utilisent une valeur de 5.

```
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586b' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315869' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c315858' rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA add disk
'/dev/mapper/3600a098038303537762b47594c31586a' rebalance power 5;
Diskgroup altered.
```

Surveiller le fonctionnement

Une opération de rééquilibrage peut être contrôlée et gérée de plusieurs manières. Nous avons utilisé la commande suivante dans cet exemple.

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

Une fois la migration terminée, aucune opération de rééquilibrage n'est signalée.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

Supprimez les anciennes LUN

La migration est maintenant terminée à mi-chemin. Il peut être souhaitable d'effectuer quelques tests de performances de base pour s'assurer que l'environnement est sain. Après confirmation, les données restantes peuvent être déplacées en déposant les anciennes LUN. Notez que cela ne provoque pas la publication immédiate des LUN. L'opération de DROP indique à Oracle ASM de déplacer d'abord les extensions, puis de libérer la LUN.

```
sqlplus / as sysasm
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0000 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup NEWDATA drop disk NEWDATA_0001 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0002 rebalance power 5;
Diskgroup altered.
SQL> alter diskgroup newdata drop disk NEWDATA_0003 rebalance power 5;
Diskgroup altered.
```

Surveiller le fonctionnement

L'opération de rééquilibrage peut être contrôlée et gérée de plusieurs manières. Nous avons utilisé la commande suivante dans cet exemple :

```
SQL> select group_number,operation,state from v$asm_operation;
GROUP_NUMBER OPERA STAT
------------ ----- ----
           1 REBAL RUN
           1 REBAL WAIT
```

Une fois la migration terminée, aucune opération de rééquilibrage n'est signalée.

```
SQL> select group_number,operation,state from v$asm_operation;
no rows selected
```

Supprimer les anciens LUN

Avant de supprimer les anciennes LUN du groupe de disques, vous devez effectuer une dernière vérification de l'état de l'en-tête. Une fois qu'une LUN est libérée d'ASM, son nom n'est plus répertorié et son état est répertorié comme `FORMER`. Cela signifie que ces LUN peuvent être supprimées du système en toute sécurité.

```
SQL> select name||' '||group_number||' '||total_mb||' '||path||'
'||header_status from v$asm_disk;
NAME||''||GROUP_NUMBER||''||TOTAL_MB||''||PATH||''||HEADER_STATUS
------------------------------------------------------------------------
------
 0 0 /dev/mapper/3600a098038303537762b47594c315863 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315864 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315861 FORMER
 0 0 /dev/mapper/3600a098038303537762b47594c315862 FORMER
NEWDATA_0005 1 10240 /dev/mapper/3600a098038303537762b47594c315869 MEMBER
NEWDATA_0007 1 10240 /dev/mapper/3600a098038303537762b47594c31586a MEMBER
NEWDATA_0004 1 10240 /dev/mapper/3600a098038303537762b47594c31586b MEMBER
NEWDATA_0006 1 10240 /dev/mapper/3600a098038303537762b47594c315858 MEMBER
8 rows selected.
```

Migration LVM

La procédure présentée ici présente les principes d'une migration basée sur LVM d'un groupe de volumes appelé `datavg`. Les exemples sont tirés du LVM Linux, mais les principes s'appliquent également à AIX, HP-UX et VxVM. Les commandes précises peuvent varier.

1. Identifiez les LUN actuellement dans le `datavg` groupe de volumes.

```
[root@host1 ~]# pvdisplay -C | grep datavg
  /dev/mapper/3600a098038303537762b47594c31582f datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31585a datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c315859 datavg lvm2 a--  10.00g
10.00g
  /dev/mapper/3600a098038303537762b47594c31586c datavg lvm2 a--  10.00g
10.00g
```

2. Créez de nouvelles LUN de taille physique identique ou légèrement supérieure et définissez-les comme volumes physiques.

```
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315864
  Physical volume "/dev/mapper/3600a098038303537762b47594c315864"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315863
  Physical volume "/dev/mapper/3600a098038303537762b47594c315863"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315862
  Physical volume "/dev/mapper/3600a098038303537762b47594c315862"
successfully created
[root@host1 ~]# pvcreate /dev/mapper/3600a098038303537762b47594c315861
  Physical volume "/dev/mapper/3600a098038303537762b47594c315861"
successfully created
```

3. Ajoutez les nouveaux volumes au groupe de volumes.

```
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315864
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315863
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315862
  Volume group "datavg" successfully extended
[root@host1 tmp]# vgextend datavg
/dev/mapper/3600a098038303537762b47594c315861
  Volume group "datavg" successfully extended
```

4. Émettez le `pvmove` Commande permettant de déplacer les extensions de chaque LUN actuelle vers la nouvelle LUN. Le `– i [seconds]` l'argument surveille la progression de l'opération.

```
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31582f
/dev/mapper/3600a098038303537762b47594c315864
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 14.2%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 28.4%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 42.5%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 57.1%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 72.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 87.3%
  /dev/mapper/3600a098038303537762b47594c31582f: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31585a
/dev/mapper/3600a098038303537762b47594c315863
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 14.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 29.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 44.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 60.1%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 75.8%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 90.9%
  /dev/mapper/3600a098038303537762b47594c31585a: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c315859
/dev/mapper/3600a098038303537762b47594c315862
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 14.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 29.8%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 45.5%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 61.1%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 76.6%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 91.7%
  /dev/mapper/3600a098038303537762b47594c315859: Moved: 100.0%
[root@host1 tmp]# pvmove -i 10
/dev/mapper/3600a098038303537762b47594c31586c
/dev/mapper/3600a098038303537762b47594c315861
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 0.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 15.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 30.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 46.0%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 61.4%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 77.2%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 92.3%
  /dev/mapper/3600a098038303537762b47594c31586c: Moved: 100.0%
```

5. Une fois ce processus terminé, supprimez les anciennes LUN du groupe de volumes à l'aide du `vgreduce` commande. En cas de réussite, la LUN peut être supprimée en toute sécurité du système.

```
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31582f
Removed "/dev/mapper/3600a098038303537762b47594c31582f" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31585a
  Removed "/dev/mapper/3600a098038303537762b47594c31585a" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c315859
  Removed "/dev/mapper/3600a098038303537762b47594c315859" from volume
group "datavg"
[root@host1 tmp]# vgreduce datavg
/dev/mapper/3600a098038303537762b47594c31586c
  Removed "/dev/mapper/3600a098038303537762b47594c31586c" from volume
group "datavg"
```

Importation de LUN étrangères

Planification

Les procédures de migration des ressources SAN à l'aide de FLI sont décrites dans NetApp "Documentation relative à l'importation de LUN étrangères de ONTAP" .

Du point de vue de la base de données et de l'hôte, aucune étape particulière n'est requise. Une fois les zones FC mises à jour et les LUN disponibles sur ONTAP, LVM doit pouvoir lire les métadonnées LVM des LUN. De plus, les groupes de volumes sont prêts à être utilisés sans étape de configuration supplémentaire. Dans de rares cas, les environnements peuvent inclure des fichiers de configuration codés en dur avec des références à la baie de stockage précédente. Par exemple, un système Linux inclus `/etc/multipath.conf` Les règles qui référençaient un WWN d'un périphérique donné doivent être mises à jour pour refléter les modifications introduites par FLI.

> Reportez-vous à la matrice de compatibilité NetApp pour plus d'informations sur les configurations prises en charge. Si votre environnement n'est pas inclus, contactez votre représentant NetApp pour obtenir de l'aide.

Cet exemple montre la migration des LUN ASM et LVM hébergées sur un serveur Linux. FLI est pris en charge par d'autres systèmes d'exploitation. Bien que les commandes côté hôte puissent différer, les principes sont les mêmes et les procédures ONTAP sont identiques.

Identifier les LUN LVM

La première étape de la préparation consiste à identifier les LUN à migrer. Dans l'exemple illustré ici, deux systèmes de fichiers SAN sont montés sur `/orabin` et `/backups`.

```
[root@host1 ~]# df -k
Filesystem                        1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root              52403200   8811464  43591736  17% /
devtmpfs                           65882776         0  65882776   0% /dev
...
fas8060-nfs-public:/install      199229440 119368128  79861312  60%
/install
/dev/mapper/sanvg-lvorabin         20961280  12348476   8612804  59%
/orabin
/dev/mapper/sanvg-lvbackups        73364480  62947536  10416944  86%
/backups
```

Le nom du groupe de volumes peut être extrait du nom du périphérique, qui utilise le format (nom du groupe de volumes)-(nom du volume logique). Dans ce cas, le groupe de volumes est appelé `sanvg`.

Le `pvdisplay` Vous pouvez utiliser la commande suivante pour identifier les LUN qui prennent en charge ce groupe de volumes. Dans ce cas, 10 LUN constituent le `sanvg` groupe de volumes.

```
[root@host1 ~]# pvdisplay -C -o pv_name,pv_size,pv_fmt,vg_name
  PV                                            PSize   VG
  /dev/mapper/3600a0980383030445424487556574266  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574267  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574268  10.00g sanvg
  /dev/mapper/3600a0980383030445424487556574269  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426a  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426b  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426c  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426d  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426e  10.00g sanvg
  /dev/mapper/3600a098038303044542448755657426f  10.00g sanvg
  /dev/sda2                                     278.38g rhel
```

Identifier les LUN ASM

Les LUN ASM doivent également être migrés. Pour obtenir le nombre de LUN et de chemins de LUN depuis sqlplus en tant qu'utilisateur sysasm, exécutez la commande suivante :

```
SQL> select path||' '||os_mb from v$asm_disk;
PATH||''||OS_MB
--------------------------------------------------------------------------
------
/dev/oracleasm/disks/ASM0 10240
/dev/oracleasm/disks/ASM9 10240
/dev/oracleasm/disks/ASM8 10240
/dev/oracleasm/disks/ASM7 10240
/dev/oracleasm/disks/ASM6 10240
/dev/oracleasm/disks/ASM5 10240
/dev/oracleasm/disks/ASM4 10240
/dev/oracleasm/disks/ASM1 10240
/dev/oracleasm/disks/ASM3 10240
/dev/oracleasm/disks/ASM2 10240
10 rows selected.
SQL>
```

Modifications du réseau FC

L'environnement actuel contient 20 LUN à migrer. Mettez à jour le SAN actuel de sorte que ONTAP puisse accéder aux LUN actuelles. Les données n'ont pas encore été migrées, mais ONTAP doit lire les informations de configuration des LUN actuelles pour créer le nouveau home pour ces données.

Au moins un port HBA sur le système AFF/FAS doit être configuré en tant que port initiateur. En outre, les zones FC doivent être mises à jour de sorte que ONTAP puisse accéder aux LUN de la baie de stockage étrangère. Certaines baies de stockage ont configuré le masquage des LUN, ce qui limite les WWN pouvant accéder à une LUN donnée. Dans ce cas, le masquage de LUN doit également être mis à jour pour autoriser l'accès aux WWN de ONTAP.

Une fois cette étape terminée, ONTAP doit être en mesure d'afficher la baie de stockage étrangère avec le `storage array show` commande. Le champ de clé renvoyé est le préfixe utilisé pour identifier la LUN étrangère sur le système. Dans l'exemple ci-dessous, les LUN de la baie étrangère `FOREIGN_1` Apparaissent dans ONTAP en utilisant le préfixe de `FOR-1`.

Identifiez le tableau étranger

```
Cluster01::> storage array show -fields name,prefix
name          prefix
------------- ------
FOREIGN_1     FOR-1
Cluster01::>
```

Identifiez les LUN étrangères

Vous pouvez lister les LUN en transmettant le `array-name` à la `storage disk show` commande. Les données renvoyées sont référencées plusieurs fois pendant la procédure de migration.

```
Cluster01::> storage disk show -array-name FOREIGN_1 -fields disk,serial
disk     serial-number
-------- -------------
FOR-1.1  800DT$HuVWBX
FOR-1.2  800DT$HuVWBZ
FOR-1.3  800DT$HuVWBW
FOR-1.4  800DT$HuVWBY
FOR-1.5  800DT$HuVWB/
FOR-1.6  800DT$HuVWBa
FOR-1.7  800DT$HuVWBd
FOR-1.8  800DT$HuVWBb
FOR-1.9  800DT$HuVWBc
FOR-1.10 800DT$HuVWBe
FOR-1.11 800DT$HuVWBf
FOR-1.12 800DT$HuVWBg
FOR-1.13 800DT$HuVWBi
FOR-1.14 800DT$HuVWBh
FOR-1.15 800DT$HuVWBj
FOR-1.16 800DT$HuVWBk
FOR-1.17 800DT$HuVWBm
FOR-1.18 800DT$HuVWBl
FOR-1.19 800DT$HuVWBo
FOR-1.20 800DT$HuVWBn
20 entries were displayed.
Cluster01::>
```

Enregistrer des LUN de baies étrangères en tant que candidats à l'importation

Les LUN étrangères sont initialement classées comme tout type de LUN particulier. Avant de pouvoir importer des données, les LUN doivent être marquées comme étrangères et par conséquent comme candidates au processus d'importation. Cette étape est terminée en transmettant le numéro de série au `storage disk modify` comme indiqué dans l'exemple suivant. Notez que ce processus balise uniquement la LUN comme étant étrangère dans ONTAP. Aucune donnée n'est écrite sur la LUN étrangère elle-même.

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign true
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign true
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign true
Cluster01::*>
```

Création de volumes pour héberger les LUN migrés

Un volume est nécessaire pour héberger les LUN migrées. La configuration exacte du volume dépend du plan global d'exploitation des fonctionnalités ONTAP. Dans cet exemple, les LUN ASM sont placées dans un volume et les LUN LVM sont placées dans un second volume. Vous pouvez ainsi gérer les LUN en tant que groupes indépendants à des fins telles que la hiérarchisation, la création de snapshots ou la définition de contrôles de QoS.

Réglez le `snapshot-policy `to `none`. Le processus de migration peut inclure une grande partie du transfert des données. Par conséquent, si des snapshots sont créés par accident, la consommation d'espace peut augmenter de façon importante, car des données indésirables sont capturées dans les snapshots.

```
Cluster01::> volume create -volume new_asm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1152] Job succeeded: Successful
Cluster01::> volume create -volume new_lvm -aggregate data_02 -size 120G
-snapshot-policy none
[Job 1153] Job succeeded: Successful
Cluster01::>
```

Créer des LUN ONTAP

Une fois les volumes créés, les nouvelles LUN doivent être créées. Normalement, la création d'une LUN nécessite que l'utilisateur indique des informations telles que la taille de LUN, mais dans ce cas, l'argument disque étranger est transmis à la commande. Par conséquent, ONTAP réplique les données de configuration actuelle du LUN à partir du numéro de série spécifié. Il utilise également la géométrie des LUN et les données de la table de partition pour ajuster l'alignement des LUN et établir des performances optimales.

Dans cette étape, les numéros de série doivent être référencés avec le tableau étranger pour s'assurer que le LUN étranger correct est associé au nouveau LUN correct.

```
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN0 -ostype
linux -foreign-disk 800DT$HuVWBW
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_asm/LUN1 -ostype
linux -foreign-disk 800DT$HuVWBX
Created a LUN of size 10g (10737418240)
...
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN8 -ostype
linux -foreign-disk 800DT$HuVWBn
Created a LUN of size 10g (10737418240)
Cluster01::*> lun create -vserver vserver1 -path /vol/new_lvm/LUN9 -ostype
linux -foreign-disk 800DT$HuVWBo
Created a LUN of size 10g (10737418240)
```

Créer des relations d'importation

Les LUN ont été créées, mais ne sont pas configurées en tant que destination de réplication. Avant de pouvoir réaliser cette étape, les LUN doivent d'abord être mises hors ligne. Cette étape supplémentaire est conçue pour protéger les données contre les erreurs de l'utilisateur. Si ONTAP permettait l'exécution d'une migration sur une LUN en ligne, une erreur typographique risquerait d'écraser les données actives. L'étape supplémentaire consistant à forcer l'utilisateur à mettre d'abord une LUN hors ligne permet de vérifier que la LUN cible correcte est utilisée comme destination de migration.

```
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN0
Warning: This command will take LUN "/vol/new_asm/LUN0" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_asm/LUN1
Warning: This command will take LUN "/vol/new_asm/LUN1" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
...
Warning: This command will take LUN "/vol/new_lvm/LUN8" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
Cluster01::*> lun offline -vserver vserver1 -path /vol/new_lvm/LUN9
Warning: This command will take LUN "/vol/new_lvm/LUN9" in Vserver
         "vserver1" offline.
Do you want to continue? {y|n}: y
```

Une fois les LUN hors ligne, vous pouvez établir la relation d'importation en transmettant le numéro de série de la LUN étrangère à `lun import create` commande.

```
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN0
-foreign-disk 800DT$HuVWBW
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_asm/LUN1
-foreign-disk 800DT$HuVWBX
...
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN8
-foreign-disk 800DT$HuVWBn
Cluster01::*> lun import create -vserver vserver1 -path /vol/new_lvm/LUN9
-foreign-disk 800DT$HuVWBo
Cluster01::*>
```

Une fois toutes les relations d'importation établies, les LUN peuvent être remis en ligne.

```
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun online -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun online -vserver vserver1 -path /vol/new_lvm/LUN9
Cluster01::*>
```

Créer le groupe initiateur

Un groupe initiateur (igroup) fait partie de l'architecture de masquage des LUN ONTAP. L'accès à une LUN nouvellement créée n'est pas accessible à moins qu'un hôte ne bénéficie au préalable d'un accès. Pour ce faire, vous devez créer un groupe initiateur qui répertorie les WWN FC ou les noms d'initiateurs iSCSI auxquels l'accès doit être accordé. Au moment de la rédaction de ce rapport, FLI était pris en charge uniquement pour les LUN FC. Cependant, la conversion en iSCSI après migration est une tâche simple, comme illustré dans la "Conversion de protocoles".

Dans cet exemple, un groupe initiateur est créé et contient deux WWN correspondant aux deux ports disponibles sur l'adaptateur HBA de l'hôte.

```
Cluster01::*> igroup create linuxhost -protocol fcp -ostype linux
-initiator 21:00:00:0e:1e:16:63:50 21:00:00:0e:1e:16:63:51
```

Mappez les nouvelles LUN sur l'hôte

Après la création du groupe initiateur, les LUN sont ensuite mappées sur le groupe initiateur défini. Ces LUN sont uniquement disponibles pour les WWN inclus dans ce groupe initiateur. NetApp suppose, à ce stade du processus de migration, que l'hôte n'a pas été segmenté vers ONTAP. Cela est important, car si l'hôte est segmenté simultanément sur la baie étrangère et le nouveau système ONTAP, il est possible de détecter sur chaque baie des LUN portant le même numéro de série. Cette situation peut entraîner des dysfonctionnements des chemins d'accès multiples ou endommager les données.

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
Cluster01::*>
```

Mise en service

Certaines perturbations lors de l'importation d'une LUN étrangère sont inévitables en raison de la nécessité de modifier la configuration du réseau FC. Cependant,

l'interruption ne doit pas durer beaucoup plus longtemps que le temps nécessaire pour redémarrer l'environnement de base de données et mettre à jour la segmentation FC pour basculer la connectivité FC de l'hôte de la LUN étrangère vers ONTAP.

Ce processus peut être résumé comme suit :

1. Mettez toutes les activités de LUN au repos sur les LUN étrangères.
2. Rediriger les connexions FC de l'hôte vers le nouveau système ONTAP.
3. Déclencher le processus d'importation.
4. Redécouvrez les LUN.
5. Redémarrez la base de données.

Inutile d'attendre la fin du processus de migration. Dès que la migration d'une LUN donnée commence, celle-ci est disponible sur ONTAP et peut assurer le service des données pendant que le processus de copie des données se poursuit. Toutes les lectures sont transmises au LUN étranger et toutes les écritures sont écrites de manière synchrone sur les deux baies. L'opération de copie est très rapide et la surcharge liée à la redirection du trafic FC est minimale. Par conséquent, tout impact sur les performances doit être transitoire et minimal. En cas de problème, vous pouvez retarder le redémarrage de l'environnement jusqu'à ce que le processus de migration soit terminé et que les relations d'importation aient été supprimées.

Arrêtez la base de données

Dans cet exemple, la première étape de la mise en veille de l'environnement consiste à arrêter la base de données.

```
[oracle@host1 bin]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base remains unchanged with value /orabin
[oracle@host1 bin]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to:
Oracle Database 12c Enterprise Edition Release 12.1.0.2.0 - 64bit
Production
With the Partitioning, Automatic Storage Management, OLAP, Advanced
Analytics
and Real Application Testing options
SQL> shutdown immediate;
Database closed.
Database dismounted.
ORACLE instance shut down.
SQL>
```

Fermez les services de grille

L'un des systèmes de fichiers SAN en cours de migration inclut également les services Oracle ASM. La mise en veille des LUN sous-jacentes nécessite la suspension des systèmes de fichiers, ce qui signifie l'arrêt des processus avec des fichiers ouverts sur ce système de fichiers.

```
[oracle@host1 bin]$ ./crsctl stop has -f
CRS-2791: Starting shutdown of Oracle High Availability Services-managed
resources on 'host1'
CRS-2673: Attempting to stop 'ora.evmd' on 'host1'
CRS-2673: Attempting to stop 'ora.DATA.dg' on 'host1'
CRS-2673: Attempting to stop 'ora.LISTENER.lsnr' on 'host1'
CRS-2677: Stop of 'ora.DATA.dg' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.asm' on 'host1'
CRS-2677: Stop of 'ora.LISTENER.lsnr' on 'host1' succeeded
CRS-2677: Stop of 'ora.evmd' on 'host1' succeeded
CRS-2677: Stop of 'ora.asm' on 'host1' succeeded
CRS-2673: Attempting to stop 'ora.cssd' on 'host1'
CRS-2677: Stop of 'ora.cssd' on 'host1' succeeded
CRS-2793: Shutdown of Oracle High Availability Services-managed resources
on 'host1' has completed
CRS-4133: Oracle High Availability Services has been stopped.
[oracle@host1 bin]$
```

Démonter les systèmes de fichiers

Si tous les processus sont arrêtés, l'opération de montage a réussi. Si l'autorisation est refusée, il doit y avoir un processus avec un verrou sur le système de fichiers. Le `fuser` permet d'identifier ces processus.

```
[root@host1 ~]# umount /orabin
[root@host1 ~]# umount /backups
```

Désactiver les groupes de volumes

Une fois tous les systèmes de fichiers d'un groupe de volumes donné démontés, le groupe de volumes peut être désactivé.

```
[root@host1 ~]# vgchange --activate n sanvg
  0 logical volume(s) in volume group "sanvg" now active
[root@host1 ~]#
```

Modifications du réseau FC

Les zones FC peuvent maintenant être mises à jour pour supprimer tout accès de l'hôte à la baie étrangère et établir l'accès à ONTAP.

Démarrer le processus d'importation

Pour démarrer les processus d'importation de LUN, exécutez `lun import start` commande.

```
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN0
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_asm/LUN1
...
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN8
Cluster01::lun import*> lun import start -vserver vserver1 -path
/vol/new_lvm/LUN9
Cluster01::lun import*>
```

Surveiller la progression de l'importation

L'opération d'importation peut être surveillée avec `lun import show` commande. Comme indiqué ci-dessous, l'importation des 20 LUN est en cours, ce qui signifie que les données sont désormais accessibles via ONTAP, même si la copie des données progresse.

```
Cluster01::lun import*> lun import show -fields path,percent-complete
vserver   foreign-disk path              percent-complete
--------- ------------ ----------------- ----------------
vserver1  800DT$HuVWB/ /vol/new_asm/LUN4 5
vserver1  800DT$HuVWBW /vol/new_asm/LUN0 5
vserver1  800DT$HuVWBX /vol/new_asm/LUN1 6
vserver1  800DT$HuVWBY /vol/new_asm/LUN2 6
vserver1  800DT$HuVWBZ /vol/new_asm/LUN3 5
vserver1  800DT$HuVWBa /vol/new_asm/LUN5 4
vserver1  800DT$HuVWBb /vol/new_asm/LUN6 4
vserver1  800DT$HuVWBc /vol/new_asm/LUN7 4
vserver1  800DT$HuVWBd /vol/new_asm/LUN8 4
vserver1  800DT$HuVWBe /vol/new_asm/LUN9 4
vserver1  800DT$HuVWBf /vol/new_lvm/LUN0 5
vserver1  800DT$HuVWBg /vol/new_lvm/LUN1 4
vserver1  800DT$HuVWBh /vol/new_lvm/LUN2 4
vserver1  800DT$HuVWBi /vol/new_lvm/LUN3 3
vserver1  800DT$HuVWBj /vol/new_lvm/LUN4 3
vserver1  800DT$HuVWBk /vol/new_lvm/LUN5 3
vserver1  800DT$HuVWBl /vol/new_lvm/LUN6 4
vserver1  800DT$HuVWBm /vol/new_lvm/LUN7 3
vserver1  800DT$HuVWBn /vol/new_lvm/LUN8 2
vserver1  800DT$HuVWBo /vol/new_lvm/LUN9 2
20 entries were displayed.
```

Si vous avez besoin d'un processus hors ligne, retardez la redécouverte ou le redémarrage des services jusqu'à ce que la `lun import show` commande indique que la migration a abouti. Vous pouvez ensuite terminer le processus de migration comme décrit à la section "Importation de LUN étrangères—fin".

Si vous avez besoin d'une migration en ligne, redécouvrez les LUN de leur nouveau domicile et accédez aux services.

Recherchez les modifications de périphérique SCSI

Dans la plupart des cas, l'option la plus simple pour redécouvrir de nouvelles LUN consiste à redémarrer l'hôte. Cela supprime automatiquement les anciens périphériques obsolètes, détecte correctement toutes les nouvelles LUN et construit les périphériques associés, tels que les périphériques multivoies. L'exemple ci-dessous montre un processus entièrement en ligne à des fins de démonstration.

Attention : avant de redémarrer un hôte, assurez-vous que toutes les entrées dans `/etc/fstab` Les ressources SAN migrées de cette référence sont commentées. Si ce n'est pas le cas et si des problèmes surviennent lors de l'accès aux LUN, le système d'exploitation risque de ne pas démarrer. Cette situation n'endommage pas les données. Cependant, il peut être très peu commode de démarrer en mode de secours ou un mode similaire et de corriger le `/etc/fstab` Afin que le système d'exploitation puisse être démarré pour permettre le dépannage.

Les LUN de la version de Linux utilisée dans cet exemple peuvent être renumérisées avec `rescan-scsi-bus.sh` commande. Si la commande réussit, chaque chemin de LUN doit apparaître dans le résultat de la commande. Le résultat de cette commande peut être difficile à interpréter, mais si la configuration de zoning et d'igroup était correcte, de nombreuses LUN doivent apparaître et inclure un `NETAPP` chaîne du fournisseur.

```
[root@host1 /]# rescan-scsi-bus.sh
Scanning SCSI subsystem for new devices
Scanning host 0 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 0 2 0 0 ...
OLD: Host: scsi0 Channel: 02 Id: 00 Lun: 00
      Vendor: LSI      Model: RAID SAS 6G 0/1  Rev: 2.13
      Type:   Direct-Access                    ANSI SCSI revision: 05
Scanning host 1 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
 Scanning for device 1 0 0 0 ...
OLD: Host: scsi1 Channel: 00 Id: 00 Lun: 00
      Vendor: Optiarc  Model: DVD RW AD-7760H  Rev: 1.41
      Type:   CD-ROM                           ANSI SCSI revision: 05
Scanning host 2 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 3 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 4 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 5 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 6 for  SCSI target IDs  0 1 2 3 4 5 6 7, all LUNs
Scanning host 7 for  all SCSI target IDs, all LUNs
 Scanning for device 7 0 0 10 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 10
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 11 ...
OLD: Host: scsi7 Channel: 00 Id: 00 Lun: 11
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 7 0 0 12 ...
...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 18
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
 Scanning for device 9 0 1 19 ...
OLD: Host: scsi9 Channel: 00 Id: 01 Lun: 19
      Vendor: NETAPP   Model: LUN C-Mode       Rev: 8300
      Type:   Direct-Access                    ANSI SCSI revision: 05
0 new or changed device(s) found.
0 remapped or resized device(s) found.
0 device(s) removed.
```

Vérifiez la présence de périphériques multivoies

Le processus de découverte des LUN déclenche également la recréation des périphériques multivoies, mais il est connu que le pilote de chemins d'accès multiples Linux présente des problèmes occasionnels. La sortie de `multipath - ll` doit être vérifié pour vérifier que la sortie semble correcte. Par exemple, le résultat ci-dessous affiche les périphériques à chemins d'accès multiples associés à un `NETAPP` chaîne du fournisseur. Chaque périphérique a quatre chemins, dont deux avec une priorité de 50 et deux avec une priorité de 10.

Bien que le résultat exact puisse varier selon les versions de Linux, ce résultat semble normal.

> Reportez-vous à la documentation des utilitaires hôtes pour connaître la version de Linux que vous utilisez pour vérifier que l' `/etc/multipath.conf` les paramètres sont corrects.

```
[root@host1 /]# multipath -ll
3600a098038303558735d493762504b36 dm-5 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:4  sdat 66:208 active ready running
| `- 9:0:1:4  sdbn 68:16  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:4  sdf  8:80   active ready running
  `- 9:0:0:4  sdz  65:144 active ready running
3600a098038303558735d493762504b2d dm-10 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:8  sdax 67:16  active ready running
| `- 9:0:1:8  sdbr 68:80  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:8  sdj  8:144  active ready running
  `- 9:0:0:8  sdad 65:208 active ready running
...
3600a098038303558735d493762504b37 dm-8 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:5  sdau 66:224 active ready running
| `- 9:0:1:5  sdbo 68:32  active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:5  sdg  8:96   active ready running
  `- 9:0:0:5  sdaa 65:160 active ready running
3600a098038303558735d493762504b4b dm-22 NETAPP  ,LUN C-Mode
size=10G features='4 queue_if_no_path pg_init_retries 50
retain_attached_hw_handle' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 7:0:1:19 sdbi 67:192 active ready running
| `- 9:0:1:19 sdcc 69:0   active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 7:0:0:19 sdu  65:64  active ready running
  `- 9:0:0:19 sdao 66:128 active ready running
```

Réactiver le groupe de volumes LVM

Si les LUN LVM ont été correctement découvertes, le système `vgchange --activate y` la commande doit réussir. C'est un bon exemple de la valeur d'un gestionnaire de volumes logiques. Une modification du WWN d'une LUN ou même d'un numéro de série n'est pas importante, car les métadonnées du groupe de volumes sont écrites sur la LUN elle-même.

Le système d'exploitation a analysé les LUN et découvert une petite quantité de données écrites sur la LUN qui l'identifie comme un volume physique appartenant au système `sanvg volumegroup`. Il a ensuite construit tous les périphériques requis. Il suffit de réactiver le groupe de volumes.

```
[root@host1 /]# vgchange --activate y sanvg
  Found duplicate PV fpCzdLTuKfy2xDZjai1NliJh3TjLUBiT: using
/dev/mapper/3600a098038303558735d493762504b46 not /dev/sdp
  Using duplicate PV /dev/mapper/3600a098038303558735d493762504b46 from
subsystem DM, ignoring /dev/sdp
  2 logical volume(s) in volume group "sanvg" now active
```

Remonter les systèmes de fichiers

Une fois le groupe de volumes réactivé, les systèmes de fichiers peuvent être montés avec toutes les données d'origine intactes. Comme nous l'avons vu précédemment, les systèmes de fichiers sont pleinement opérationnels, même si la réplication des données est toujours active dans le groupe en arrière-plan.

```
[root@host1 /]# mount /orabin
[root@host1 /]# mount /backups
[root@host1 /]# df -k
Filesystem                        1K-blocks      Used Available Use%
Mounted on
/dev/mapper/rhel-root              52403200   8837100  43566100  17% /
devtmpfs                           65882776         0  65882776   0% /dev
tmpfs                               6291456        84   6291372   1%
/dev/shm
tmpfs                              65898668      9884  65888784   1% /run
tmpfs                              65898668         0  65898668   0%
/sys/fs/cgroup
/dev/sda1                            505580    224828    280752  45% /boot
fas8060-nfs-public:/install      199229440 119368256  79861184  60%
/install
fas8040-nfs-routable:/snapomatic   9961472     30528   9930944   1%
/snapomatic
tmpfs                              13179736        16  13179720   1%
/run/user/42
tmpfs                              13179736         0  13179736   0%
/run/user/0
/dev/mapper/sanvg-lvorabin         20961280  12357456   8603824  59%
/orabin
/dev/mapper/sanvg-lvbackups        73364480  62947536  10416944  86%
/backups
```

Rechercher à nouveau les périphériques ASM

Les périphériques ASMlib auraient dû être redécouverts lorsque les périphériques SCSI ont été renumérisés. La redécouverte peut être vérifiée en ligne en redémarrant ASMlib puis en analysant les disques.

> Cette étape concerne uniquement les configurations ASM où ASMlib est utilisé.

Attention : lorsque ASMlib n'est pas utilisé, le `/dev/mapper` les périphériques doivent avoir été recréés automatiquement. Cependant, les autorisations peuvent ne pas être correctes. Vous devez définir des autorisations spéciales sur les périphériques sous-jacents pour ASM en l'absence d'ASMlib. Cette opération est généralement réalisée par des entrées spéciales dans l'un ou l'autre des `/etc/multipath.conf` ou `udev` ou éventuellement dans les deux jeux de règles. Ces fichiers peuvent avoir besoin d'être mis à jour pour refléter les modifications de l'environnement en termes de WWN ou de numéros de série afin de s'assurer que les périphériques ASM disposent toujours des autorisations appropriées.

Dans cet exemple, le redémarrage d'ASMlib et l'analyse des disques affichent les 10 mêmes LUN ASM que l'environnement d'origine.

```
[root@host1 /]# oracleasm exit
Unmounting ASMlib driver filesystem: /dev/oracleasm
Unloading module "oracleasm": oracleasm
[root@host1 /]# oracleasm init
Loading module "oracleasm": oracleasm
Configuring "oracleasm" to use device physical block size
Mounting ASMlib driver filesystem: /dev/oracleasm
[root@host1 /]# oracleasm scandisks
Reloading disk partitions: done
Cleaning any stale ASM disks...
Scanning system for ASM disks...
Instantiating disk "ASM0"
Instantiating disk "ASM1"
Instantiating disk "ASM2"
Instantiating disk "ASM3"
Instantiating disk "ASM4"
Instantiating disk "ASM5"
Instantiating disk "ASM6"
Instantiating disk "ASM7"
Instantiating disk "ASM8"
Instantiating disk "ASM9"
```

Redémarrez les services de grille

Maintenant que les périphériques LVM et ASM sont en ligne et disponibles, les services de grille peuvent être redémarrés.

```
[root@host1 /]# cd /orabin/product/12.1.0/grid/bin
[root@host1 bin]# ./crsctl start has
```

Redémarrez la base de données

Une fois les services de grille redémarrés, la base de données peut être ouverte. Il peut être nécessaire d’attendre quelques minutes que les services ASM soient entièrement disponibles avant d’essayer de démarrer la base de données.

```
[root@host1 bin]# su - oracle
[oracle@host1 ~]$ . oraenv
ORACLE_SID = [oracle] ? FLIDB
The Oracle base has been set to /orabin
[oracle@host1 ~]$ sqlplus / as sysdba
SQL*Plus: Release 12.1.0.2.0
Copyright (c) 1982, 2014, Oracle.  All rights reserved.
Connected to an idle instance.
SQL> startup
ORACLE instance started.
Total System Global Area 3221225472 bytes
Fixed Size                  4502416 bytes
Variable Size            1207962736 bytes
Database Buffers         1996488704 bytes
Redo Buffers               12271616 bytes
Database mounted.
Database opened.
SQL>
```

Achèvement

Du point de vue de l’hôte, la migration est terminée, mais les E/S sont toujours servies depuis la baie étrangère jusqu’à ce que les relations d’importation soient supprimées.

Avant de supprimer les relations, vous devez confirmer que le processus de migration est terminé pour toutes les LUN.

```
Cluster01::*> lun import show -vserver vserver1 -fields foreign-
disk,path,operational-state
vserver   foreign-disk path              operational-state
--------- ------------ ----------------- -----------------
vserver1 800DT$HuVWB/ /vol/new_asm/LUN4 completed
vserver1 800DT$HuVWBW /vol/new_asm/LUN0 completed
vserver1 800DT$HuVWBX /vol/new_asm/LUN1 completed
vserver1 800DT$HuVWBY /vol/new_asm/LUN2 completed
vserver1 800DT$HuVWBZ /vol/new_asm/LUN3 completed
vserver1 800DT$HuVWBa /vol/new_asm/LUN5 completed
vserver1 800DT$HuVWBb /vol/new_asm/LUN6 completed
vserver1 800DT$HuVWBc /vol/new_asm/LUN7 completed
vserver1 800DT$HuVWBd /vol/new_asm/LUN8 completed
vserver1 800DT$HuVWBe /vol/new_asm/LUN9 completed
vserver1 800DT$HuVWBf /vol/new_lvm/LUN0 completed
vserver1 800DT$HuVWBg /vol/new_lvm/LUN1 completed
vserver1 800DT$HuVWBh /vol/new_lvm/LUN2 completed
vserver1 800DT$HuVWBi /vol/new_lvm/LUN3 completed
vserver1 800DT$HuVWBj /vol/new_lvm/LUN4 completed
vserver1 800DT$HuVWBk /vol/new_lvm/LUN5 completed
vserver1 800DT$HuVWBl /vol/new_lvm/LUN6 completed
vserver1 800DT$HuVWBm /vol/new_lvm/LUN7 completed
vserver1 800DT$HuVWBn /vol/new_lvm/LUN8 completed
vserver1 800DT$HuVWBo /vol/new_lvm/LUN9 completed
20 entries were displayed.
```

Supprimer les relations d'importation

Une fois le processus de migration terminé, supprimez la relation de migration. Une fois que vous avez terminé, les E/S sont servies exclusivement à partir des disques sur ONTAP.

```
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN0
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_asm/LUN1
...
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN8
Cluster01::*> lun import delete -vserver vserver1 -path /vol/new_lvm/LUN9
```

Désenregistrer des LUN étrangères

Enfin, modifiez le disque pour retirer le `is-foreign` désignation.

```
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBW} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBX} -is
-foreign false
...
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBn} -is
-foreign false
Cluster01::*> storage disk modify {-serial-number 800DT$HuVWBo} -is
-foreign false
Cluster01::*>
```

Conversion de protocoles

La modification du protocole utilisé pour accéder à une LUN est une exigence courante.

Dans certains cas, cela fait partie d'une stratégie globale de migration des données vers le cloud. Le protocole TCP/IP est le protocole du cloud. En passant de FC à iSCSI, vous simplifiez la migration vers divers environnements cloud. Dans d'autres cas, il peut être souhaitable de tirer parti de la réduction des coûts d'un SAN IP. Il arrive qu'une migration utilise un protocole différent comme mesure temporaire. Par exemple, si une baie étrangère et des LUN ONTAP ne peuvent pas coexister sur les mêmes HBA, vous pouvez utiliser des LUN iSCSI suffisamment longues pour copier les données de l'ancienne baie. Vous pouvez ensuite reconvertir en FC après le retrait des anciennes LUN du système.

La procédure suivante illustre la conversion de FC en iSCSI, mais les principes généraux s'appliquent à une conversion iSCSI inverse en FC.

Installez l'initiateur iSCSI

La plupart des systèmes d'exploitation incluent par défaut un initiateur iSCSI logiciel, mais si celui-ci n'est pas inclus, il peut être facilement installé.

```
[root@host1 /]# yum install -y iscsi-initiator-utils
Loaded plugins: langpacks, product-id, search-disabled-repos,
subscription-
              : manager
Resolving Dependencies
--> Running transaction check
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.el7 will be
updated
--> Processing Dependency: iscsi-initiator-utils = 6.2.0.873-32.el7 for
package: iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
---> Package iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7 will be
an update
--> Running transaction check
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.el7 will
be updated
---> Package iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
```

```
will be an update
--> Finished Dependency Resolution
Dependencies Resolved
================================================================================
===
 Package                        Arch   Version              Repository
Size
================================================================================
===
Updating:
 iscsi-initiator-utils          x86_64 6.2.0.873-32.0.2.el7 ol7_latest 416
k
Updating for dependencies:
 iscsi-initiator-utils-iscsiuio x86_64 6.2.0.873-32.0.2.el7 ol7_latest  84
k
Transaction Summary
================================================================================
===
Upgrade  1 Package (+1 Dependent package)
Total download size: 501 k
Downloading packages:
No Presto metadata available for ol7_latest
(1/2): iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_6 | 416 kB   00:00
(2/2): iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2. |  84 kB   00:00
--------------------------------------------------------------------------------
---
Total                                           2.8 MB/s | 501 kB
00:00Cluster01
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
  Updating   : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
1/4
  Updating   : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
2/4
  Cleanup    : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
3/4
  Cleanup    : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
rhel-7-server-eus-rpms/7Server/x86_64/productid       | 1.7 kB     00:00
rhel-7-server-rpms/7Server/x86_64/productid           | 1.7 kB     00:00
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.0.2.el7.x86_64
1/4
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.0.2.el7.x86
2/4
```

```
  Verifying  : iscsi-initiator-utils-iscsiuio-6.2.0.873-32.el7.x86_64
3/4
  Verifying  : iscsi-initiator-utils-6.2.0.873-32.el7.x86_64
4/4
Updated:
  iscsi-initiator-utils.x86_64 0:6.2.0.873-32.0.2.el7
Dependency Updated:
  iscsi-initiator-utils-iscsiuio.x86_64 0:6.2.0.873-32.0.2.el7
Complete!
[root@host1 /]#
```

Identifiez le nom de l'initiateur iSCSI

Un nom d'initiateur iSCSI unique est généré lors du processus d'installation. Sous Linux, il se trouve dans le `/etc/iscsi/initiatorname.iscsi` fichier. Ce nom permet d'identifier l'hôte sur le SAN IP.

```
[root@host1 /]# cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1992-05.com.redhat:497bd66ca0
```

Créer un nouveau groupe initiateur

Un groupe initiateur (igroup) fait partie de l'architecture de masquage des LUN ONTAP. L'accès à une LUN nouvellement créée n'est pas accessible à moins qu'un hôte ne bénéficie au préalable d'un accès. Cette étape est effectuée en créant un groupe initiateur qui répertorie les WWN FC ou les noms d'initiateurs iSCSI nécessitant un accès.

Dans cet exemple, un groupe initiateur contenant l'initiateur iSCSI de l'hôte Linux est créé.

```
Cluster01::*> igroup create -igroup linuxiscsi -protocol iscsi -ostype
linux -initiator iqn.1994-05.com.redhat:497bd66ca0
```

Arrêtez l'environnement

Avant de modifier le protocole LUN, les LUN doivent être complètement suspendues. Toute base de données de l'une des LUN en cours de conversion doit être arrêtée, les systèmes de fichiers doivent être démontés et les groupes de volumes doivent être désactivés. Si ASM est utilisé, assurez-vous que le groupe de disques ASM est démonté et arrêtez tous les services de grille.

Annulez le mappage des LUN à partir du réseau FC

Une fois les LUN entièrement suspendues, supprimez les mappages du groupe initiateur FC d'origine.

```
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxhost
...
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxhost
Cluster01::*> lun unmap -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxhost
```

Remappez les LUN sur le réseau IP

Accordez l'accès à chaque LUN au nouveau groupe initiateur iSCSI.

```
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN0 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_asm/LUN1 -igroup
linuxiscsi
...
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN8 -igroup
linuxiscsi
Cluster01::*> lun map -vserver vserver1 -path /vol/new_lvm/LUN9 -igroup
linuxiscsi
Cluster01::*>
```

Découvrez les cibles iSCSI

La découverte iSCSI se déroule en deux phases. Le premier consiste à découvrir les cibles, qui n'équivaut pas à détecter une LUN. Le `iscsiadm` la commande illustrée ci-dessous sonde le groupe de portails spécifié par le `-p argument` Et stocke une liste de toutes les adresses IP et de tous les ports qui offrent des services iSCSI. Dans ce cas, quatre adresses IP disposent de services iSCSI sur le port par défaut 3260.

> Cette commande peut prendre plusieurs minutes si l'une des adresses IP cibles ne peut pas être atteinte.

```
[root@host1 ~]# iscsiadm -m discovery -t st -p fas8060-iscsi-public1
10.63.147.197:3260,1033 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
10.63.147.198:3260,1034 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.203:3260,1030 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
172.20.108.202:3260,1029 iqn.1992-
08.com.netapp:sn.807615e9ef6111e5a5ae90e2ba5b9464:vs.3
```

Découverte des LUN iSCSI

Une fois les cibles iSCSI détectées, redémarrez le service iSCSI pour découvrir les LUN iSCSI disponibles et construire les périphériques associés tels que les périphériques multivoies ou ASMlib.

```
[root@host1 ~]# service iscsi restart
Redirecting to /bin/systemctl restart  iscsi.service
```

Redémarrez l'environnement

Redémarrez l'environnement en réactivant les groupes de volumes, en remontant les systèmes de fichiers, en redémarrant les services RAC, etc. Par mesure de précaution, NetApp vous recommande de redémarrer le serveur une fois le processus de conversion terminé afin de vous assurer que tous les fichiers de configuration sont corrects et que tous les périphériques obsolètes sont supprimés.

Attention : avant de redémarrer un hôte, assurez-vous que toutes les entrées dans `/etc/fstab` Les ressources SAN migrées de cette référence sont commentées. Si cette étape n'est pas effectuée et qu'il y a des problèmes avec l'accès aux LUN, le système d'exploitation ne s'amorce pas. Ce problème n'endommage pas les données. Cependant, il peut être très peu commode de démarrer en mode de secours ou un mode similaire et correct `/etc/fstab` Afin que le système d'exploitation puisse être démarré pour permettre aux efforts de dépannage de commencer.

Exemples de scripts

Les scripts présentés sont fournis sous forme d'exemples de script de diverses tâches du système d'exploitation et de la base de données. Ils sont fournis en l'état. Si une assistance est requise pour une procédure particulière, contactez NetApp ou un revendeur NetApp.

Arrêt de la base de données

Le script Perl suivant prend un seul argument du SID Oracle et arrête une base de données. Il peut être exécuté en tant qu'utilisateur Oracle ou en tant que root.

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
77 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF4
sqlplus / as sysdba << EOF2
shutdown immediate;
EOF2
`;};
print @out;
if ("@out" =~ /ORACLE instance shut down/) {
print "$oraclesid shut down\n";
exit 0;}
elsif ("@out" =~ /Connected to an idle instance/) {
print "$oraclesid already shut down\n";
exit 0;}
else {
print "$oraclesid failed to shut down\n";
exit 1;}
```

Démarrage de la base de données

Le script Perl suivant prend un seul argument du SID Oracle et arrête une base de données. Il peut être exécuté en tant qu'utilisateur Oracle ou en tant que root.

```
#! /usr/bin/perl
use strict;
use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
my $uid=$<;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
'
`;}
else {
@out=`. oraenv << EOF3
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
startup;
EOF2
`;};
print @out;
if ("@out" =~ /Database opened/) {
print "$oraclesid started\n";
exit 0;}
elsif ("@out" =~ /cannot start already-running ORACLE/) {
print "$oraclesid already started\n";
exit 1;}
else {
78 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
print "$oraclesid failed to start\n";
exit 1;}
```

Convertir le système de fichiers en lecture seule

Le script suivant prend un argument de système de fichiers et tente de le démonter et de le remonter en lecture seule. Cette opération est utile lors des processus de migration au cours desquels un système de fichiers doit être mis à disposition pour répliquer des données, tout en étant protégé contre les dommages accidentels.

```
#! /usr/bin/perl
use strict;
#use warnings;
my $filesystem=$ARGV[0];
my @out=`umount '$filesystem'`;
if ($? == 0) {
 print "$filesystem unmounted\n";
 @out = `mount -o ro '$filesystem'`;
 if ($? == 0) {
 print "$filesystem mounted read-only\n";
 exit 0;}}
else {
 print "Unable to unmount $filesystem\n";
 exit 1;}
print @out;
```

Remplacer le système de fichiers

L'exemple de script suivant est utilisé pour remplacer un système de fichiers par un autre. Comme il modifie le fichier `/etc/fstab `file, il doit être exécuté en tant que root. Il accepte un seul argument délimité par des virgules pour les anciens et les nouveaux systèmes de fichiers.

1. Pour remplacer le système de fichiers, exécutez le script suivant :

```
#! /usr/bin/perl
use strict;
#use warnings;
my $oldfs;
my $newfs;
my @oldfstab;
my @newfstab;
my $source;
my $mountpoint;
my $leftover;
my $oldfstabentry='';
my $newfstabentry='';
my $migratedfstabentry='';
($oldfs, $newfs) = split (',',$ARGV[0]);
open(my $filehandle, '<', '/etc/fstab') or die "Could not open
/etc/fstab\n";
while (my $line = <$filehandle>) {
 chomp $line;
 ($source, $mountpoint, $leftover) = split(/[ , ]/,$line, 3);
 if ($mountpoint eq $oldfs) {
 $oldfstabentry = "#Removed by swap script $source $oldfs $leftover";}
```

```
 elsif ($mountpoint eq $newfs) {
 $newfstabentry = "#Removed by swap script $source $newfs $leftover";
 $migratedfstabentry = "$source $oldfs $leftover";}
 else {
 push (@newfstab, "$line\n")}}
79 Migration of Oracle Databases to NetApp Storage Systems © 2021
NetApp, Inc. All rights reserved
push (@newfstab, "$oldfstabentry\n");
push (@newfstab, "$newfstabentry\n");
push (@newfstab, "$migratedfstabentry\n");
close($filehandle);
if ($oldfstabentry eq ''){
 die "Could not find $oldfs in /etc/fstab\n";}
if ($newfstabentry eq ''){
 die "Could not find $newfs in /etc/fstab\n";}
my @out=`umount '$newfs'`;
if ($? == 0) {
 print "$newfs unmounted\n";}
else {
 print "Unable to unmount $newfs\n";
 exit 1;}
@out=`umount '$oldfs'`;
if ($? == 0) {
 print "$oldfs unmounted\n";}
else {
 print "Unable to unmount $oldfs\n";
 exit 1;}
system("cp /etc/fstab /etc/fstab.bak");
open ($filehandle, ">", '/etc/fstab') or die "Could not open /etc/fstab
for writing\n";
for my $line (@newfstab) {
 print $filehandle $line;}
close($filehandle);
@out=`mount '$oldfs'`;
if ($? == 0) {
 print "Mounted updated $oldfs\n";
 exit 0;}
else{
 print "Unable to mount updated $oldfs\n";
 exit 1;}
exit 0;
```

Comme exemple d'utilisation de ce script, supposons que les données dans `/oradata` est migré vers `/neworadata` et `/logs` est migré vers `/newlogs`. L'une des méthodes les plus simples pour effectuer cette tâche consiste à utiliser une simple opération de copie de fichier pour replacer le nouveau périphérique sur le point de montage d'origine.

2. Supposons que l'ancien et le nouveau système de fichiers sont présents dans le `/etc/fstab` classer comme suit :

```
cluster01:/vol_oradata /oradata nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
cluster01:/vol_logs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /newlogs nfs rw,bg,vers=3,rsize=65536,wsize=65536
0 0
```

3. Lors de son exécution, ce script démonte le système de fichiers actuel et le remplace par le nouveau :

```
[root@jfsc3 scripts]# ./swap.fs.pl /oradata,/neworadata
/neworadata unmounted
/oradata unmounted
Mounted updated /oradata
[root@jfsc3 scripts]# ./swap.fs.pl /logs,/newlogs
/newlogs unmounted
/logs unmounted
Mounted updated /logs
```

4. Le script met également à jour le `/etc/fstab` classez-les en conséquence. Dans l'exemple illustré ici, il inclut les modifications suivantes :

```
#Removed by swap script cluster01:/vol_oradata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_neworadata /neworadata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_neworadata /oradata nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_logs /logs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
#Removed by swap script cluster01:/vol_newlogs /newlogs nfs
rw,bg,vers=3,rsize=65536,wsize=65536 0 0
cluster01:/vol_newlogs /logs nfs rw,bg,vers=3,rsize=65536,wsize=65536 0
0
```

Migration automatisée des bases de données

Cet exemple illustre l'utilisation de scripts d'arrêt, de démarrage et de remplacement de système de fichiers pour automatiser complètement la migration.

```
#! /usr/bin/perl
use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my @oldfs;
my @newfs;
my $x=1;
while ($x < scalar(@ARGV)) {
   ($oldfs[$x-1], $newfs[$x-1]) = split (',',$ARGV[$x]);
   $x+=1;}
my @out=`./dbshut.pl '$oraclesid'`;
print @out;
if ($? ne 0) {
   print "Failed to shut down database\n";
   exit 0;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`./mk.fs.readonly.pl '$oldfs[$x]'`;
   if ($? ne 0) {
      print "Failed to make filesystem $oldfs[$x] readonly\n";
      exit 0;}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   my @out=`rsync -rlpogt --stats --progress --exclude='.snapshot'
'$oldfs[$x]/' '/$newfs[$x]/'`;
   print @out;
   if ($? ne 0) {
      print "Failed to copy filesystem $oldfs[$x] to $newfs[$x]\n";
      exit 0;}
   else {
      print "Succesfully replicated filesystem $oldfs[$x] to
$newfs[$x]\n";}
   $x+=1;}
$x=0;
while ($x < scalar(@oldfs)) {
   print "swap $x $oldfs[$x] $newfs[$x]\n";
   my @out=`./swap.fs.pl '$oldfs[$x],$newfs[$x]'`;
   print @out;
   if ($? ne 0) {
      print "Failed to swap filesystem $oldfs[$x] for $newfs[$x]\n";
      exit 1;}
   else {
      print "Swapped filesystem $oldfs[$x] for $newfs[$x]\n";}
   $x+=1;}
my @out=`./dbstart.pl '$oraclesid'`;
```

```
print @out;
```

Afficher les emplacements des fichiers

Ce script collecte un certain nombre de paramètres de base de données critiques et les imprime dans un format facile à lire. Ce script peut être utile lors de la révision des dispositions de données. En outre, le script peut être modifié pour d'autres utilisations.

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;};
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print "$oraclesid datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
```

```
@out=dosql('select member from v\\\\\$logfile;');
print "$oraclesid redo logs:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name from v\\\\\$tempfile;');
print "$oraclesid temp datafiles:\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('show parameter spfile;');
print "$oraclesid spfile\n";
for $line (@out) {
        chomp($line);
        if (length($line)>0) {print "$line\n";}}
print "\n";
@out=dosql('select name||\' \'||value from v\\\\\$parameter where
isdefault=\'FALSE\';');
print "$oraclesid key parameters\n";
for $line (@out) {
        chomp($line);
        if ($line =~ /control_files/) {print "$line\n";}
        if ($line =~ /db_create/) {print "$line\n";}
        if ($line =~ /db_file_name_convert/) {print "$line\n";}
        if ($line =~ /log_archive_dest/) {print "$line\n";}}
        if ($line =~ /log_file_name_convert/) {print "$line\n";}
        if ($line =~ /pdb_file_name_convert/) {print "$line\n";}
        if ($line =~ /spfile/) {print "$line\n";}
print "\n";
```

Nettoyage de la migration ASM

```
#! /usr/bin/perl
#use strict;
#use warnings;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my @out;
sub dosql{
        my $command = @_[0];
        my @lines;
        my $uid=$<;
        if ($uid == 0) {
```

```
        @lines=`su - $oracleuser -c "export ORAENV_ASK=NO;export
ORACLE_SID=$oraclesid;. oraenv -s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
"
        `;}
        else {
        $command=~s/\\\\\\/\\/g;
        @lines=`export ORAENV_ASK=NO;export ORACLE_SID=$oraclesid;. oraenv
-s << EOF1
EOF1
sqlplus -S / as sysdba << EOF2
set heading off
$command
EOF2
        `;}
return @lines}
print "\n";
@out=dosql('select name from v\\\\\$datafile;');
print @out;
print "shutdown immediate;\n";
print "startup mount;\n";
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "host mv $line $1$newname\n";}}
print "\n";
for $line (@out) {
        if (length($line) > 1) {
                chomp($line);
                ($first, $second,$third,$fourth)=split('_',$line);
                $fourth =~ s/^TS-//;
                $newname=lc("$fourth.dbf");
                $path2file=$line;
                $path2file=~ /(^.*.\/)/;
                print "alter database rename file '$line' to
'$1$newname';\n";}}
```

```
print "alter database open;\n";
print "\n";
```

Conversion du nom ASM en nom de système de fichiers

```
set serveroutput on;
set wrap off;
declare
    cursor df is select file#, name from v$datafile;
    cursor tf is select file#, name from v$tempfile;
    cursor lf is select member from v$logfile;
    firstline boolean := true;
begin
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('Parameters for log file conversion:');
    dbms_output.put_line(CHR(13));
    dbms_output.put('*.log_file_name_convert = ');
    for lfrec in lf loop
        if (firstline = true) then
            dbms_output.put('''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*./','') || '''');
        else
            dbms_output.put(',''' || lfrec.member || ''', ');
            dbms_output.put('''/NEW_PATH/' ||
regexp_replace(lfrec.member,'^.*./','') || '''');
        end if;
        firstline:=false;
    end loop;
    dbms_output.put_line(CHR(13));
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('rman duplication script:');
    dbms_output.put_line(CHR(13));
    dbms_output.put_line('run');
    dbms_output.put_line('{');
    for dfrec in df loop
        dbms_output.put_line('set newname for datafile ' ||
            dfrec.file# || ' to ''' || dfrec.name ||''';');
    end loop;
    for tfrec in tf loop
        dbms_output.put_line('set newname for tempfile ' ||
            tfrec.file# || ' to ''' || tfrec.name ||''';');
    end loop;
    dbms_output.put_line('duplicate target database for standby backup
location INSERT_PATH_HERE;');
    dbms_output.put_line('}');
end;
/
```

Relire les journaux sur la base de données

Ce script accepte un seul argument d'un SID Oracle pour une base de données en mode montage et tente de relire tous les journaux d'archives actuellement disponibles.

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
84 Migration of Oracle Databases to NetApp Storage Systems © 2021 NetApp,
Inc. All rights reserved
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover database until cancel;
auto
EOF2
`;
}
print @out;
```

Relire les journaux sur la base de données de secours

Ce script est identique au script précédent, sauf qu'il est conçu pour une base de données de secours.

```
#! /usr/bin/perl
use strict;
my $oraclesid=$ARGV[0];
my $oracleuser='oracle';
my $uid = $<;
my @out;
if ($uid == 0) {
@out=`su - $oracleuser -c '. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
'
`;}
else {
@out=`. oraenv << EOF1
$oraclesid
EOF1
sqlplus / as sysdba << EOF2
recover standby database until cancel;
auto
EOF2
`;
}
print @out;
```

Remarques supplémentaires

Optimisation des performances et analyse comparative

Il est extrêmement compliqué de tester précisément les performances du stockage des bases de données. Il faut comprendre les problèmes suivants :

- IOPS et débit
- Différence entre les opérations d'E/S au premier plan et en arrière-plan
- Effet de la latence sur la base de données
- Nombreux paramètres du système d'exploitation et du réseau qui affectent également les performances du stockage

En outre, il faut tenir compte des tâches qui ne relèvent pas du domaine du stockage dans les bases de données. L'optimisation de la performance du stockage ne présente plus d'avantages, car la performance du stockage n'est plus un facteur limitant.

La majorité des clients de base de données choisissent désormais des baies 100 % Flash, ce qui entraîne d'autres considérations. Prenons l'exemple des tests de performances sur un système AFF A900 à deux nœuds :

- Avec un ratio de lecture/écriture de 80/20, deux nœuds A900 peuvent fournir plus de 1 million d'IOPS de base de données aléatoires avant que la latence ne dépasse même le seuil de 150 µs. Au-delà des exigences de performance actuelles de la plupart des bases de données, il est difficile de prévoir l'amélioration attendue. Le stockage serait largement effacé comme un goulot d'étranglement.
- La bande passante réseau est une source de plus en plus courante de limites de performances. Par exemple, les solutions sur disque mécanique constituent souvent des goulots d'étranglement pour les performances des bases de données, car la latence d'E/S est très élevée. Lorsque les limites de latence sont éliminées par un système 100 % Flash, le obstacle est fréquemment basculer vers le réseau. Ceci est particulièrement notable dans les environnements virtualisés et les systèmes lames où la véritable connectivité réseau est difficile à visualiser. Les tests de performances peuvent ainsi être plus complexes si le système de stockage lui-même ne peut pas être pleinement utilisé en raison des limitations de bande passante.
- Il est généralement impossible de comparer les performances d'une baie 100 % Flash à celles d'une baie contenant des disques rotatifs en raison de la latence considérablement améliorée des baies 100 % Flash. Les résultats des tests ne sont généralement pas significatifs.
- Généralement, comparer les pics de performance d'IOPS avec un système 100 % Flash n'est pas utile, car les bases de données ne sont pas limitées par les E/S de stockage Supposons par exemple qu'une baie peut supporter 500 000 IOPS aléatoires, tandis qu'une autre peut supporter 300 000. La différence n'est pas pertinente en situation réelle si une base de données consacre 99 % de son temps au traitement du processeur. Ces charges de travail n'exploitent jamais toutes les capacités de la baie de stockage. À l'inverse, les pics d'activité d'E/S par seconde peuvent s'avérer critiques pour une plateforme de consolidation sur laquelle la baie de stockage doit être chargée au maximum de ses capacités.
- Lors de tout test de stockage, la latence et les IOPS sont systématiquement prises en compte. De nombreuses baies de stockage sur le marché revendiquant des niveaux extrêmes d'IOPS, mais avec la latence, ces IOPS deviennent inutiles à de tels niveaux. La cible type avec des baies 100 % Flash est le millième de seconde, Une meilleure approche lors de ces tests n'est pas de mesurer les IOPS maximales mais de déterminer le nombre d'IOPS qu'une baie de stockage peut supporter avant que la latence moyenne ne soit supérieure à 1 ms.

Référentiel automatique de workloads Oracle et banc d'essai

Pour les comparaisons de performances Oracle, il est référence dans le rapport Oracle Automatic Workload Repository (AWR).

Il existe plusieurs types de rapports AWR. Du point de vue du stockage, un rapport généré par l'exécution de `awrrpt.sql` La commande est la plus complète et la plus utile, car elle cible une instance de base de données spécifique et inclut des histogrammes détaillés qui décomposent les événements d'E/S de stockage en fonction de la latence.

Dans l'idéal, comparer deux baies de performances implique d'exécuter la même charge de travail sur chaque baie et de produire un rapport AWR qui cible précisément la charge de travail. Dans le cas d'une charge de travail très longue, il est possible d'utiliser un seul rapport AWR avec un temps écoulé couvrant le temps de début et de fin, mais il est préférable de séparer les données AWR sous forme de plusieurs rapports. Par exemple, si une tâche par lots s'est exécutée de minuit à 6 h, créez une série de rapports AWR d'une heure de minuit à 1 h, de 1 h à 2 h, etc.

Dans d'autres cas, une requête très courte doit être optimisée. La meilleure option est un rapport AWR basé sur un instantané AWR créé au début de la requête et un deuxième instantané AWR créé à la fin de la requête. Le serveur de base de données doit être silencieux pour réduire au minimum l'activité en arrière-plan

qui pourrait masquer l'activité de la requête en cours d'analyse.

> Lorsque les rapports AWR ne sont pas disponibles, les rapports Oracle statspack constituent une bonne alternative. Ils contiennent la plupart des mêmes statistiques d'E/S qu'un rapport AWR.

Oracle AWR et dépannage

Un rapport AWR est également l'outil le plus important pour analyser un problème de performances.

Comme pour les bancs d'essai, la résolution des problèmes de performances nécessite que vous mesuriez précisément une charge de travail particulière. Dans la mesure du possible, fournissez des données AWR lorsque vous signalez un problème de performance au centre de support NetApp ou lorsque vous travaillez avec une équipe NetApp ou un partenaire responsable de compte concernant une nouvelle solution.

Lorsque vous fournissez des données AWR, tenez compte des exigences suivantes :

- Exécutez le `awrrpt.sql` pour générer le rapport. La sortie peut être texte ou HTML.
- Si Oracle Real application clusters (RAC) est utilisé, générez des rapports AWR pour chaque instance du cluster.
- Cibler l'heure précise à laquelle le problème a existé. La durée maximale acceptable d'un rapport AWR est généralement d'une heure. Si un problème persiste pendant plusieurs heures ou implique une opération sur plusieurs heures, par exemple un traitement par lots, fournissez plusieurs rapports AWR d'une heure qui couvrent l'ensemble de la période à analyser.
- Si possible, réglez l'intervalle d'instantané AWR sur 15 minutes. Ce paramètre permet d'effectuer une analyse plus détaillée. Cela nécessite également des exécutions supplémentaires de `awrrpt.sql` fournir un rapport pour chaque intervalle de 15 minutes.
- Si le problème est une requête en cours très courte, fournissez un rapport AWR basé sur un instantané AWR créé au début de l'opération et un second instantané AWR créé à la fin de l'opération. Le serveur de base de données doit être silencieux pour minimiser l'activité en arrière-plan qui pourrait masquer l'activité de l'opération en cours d'analyse.
- Si un problème de performance est signalé à certains moments mais pas à d'autres, fournissez des données AWR supplémentaires qui démontrent de bonnes performances pour la comparaison.

etalonnez_io

Le `calibrate_io` command ne doit jamais être utilisé pour tester, comparer ou tester les systèmes de stockage. Comme indiqué dans la documentation Oracle, cette procédure permet d'étalonner les capacités d'E/S du stockage.

L'étalonnage n'est pas le même que l'étalonnage. L'objectif de cette commande est d'émettre des E/S pour aider à étalonner les opérations de la base de données et améliorer leur efficacité en optimisant le niveau d'E/S émis pour l'hôte. Car le type d'E/S effectué par le `calibrate_io` Le fonctionnement ne représente pas les E/S réelles de l'utilisateur de la base de données, les résultats ne sont pas prévisibles et ne sont souvent même pas reproductibles.

SLOB2

SLOB2, le très petit banc d'essai Oracle, est devenu l'outil privilégié pour évaluer les performances des bases de données. Il a été développé par Kevin Closson et est disponible à l'adresse "https://kevinclosson.net/slob/". L'installation et la configuration ne prennent que quelques minutes et une base de données Oracle génère des modèles d'E/S sur un espace de table définissable par l'utilisateur. Il s'agit de l'une des rares options de test

disponibles permettant de saturer une baie 100 % Flash par E/S. Il est également utile de générer des niveaux d'E/S beaucoup plus bas pour simuler des charges de travail de stockage qui font partie des IOPS faibles, mais qui sont sensibles à la latence.

Swingbench

Swingbench peut être utile pour tester les performances des bases de données, mais il est extrêmement difficile d'utiliser Swingbench sous un contrainte de stockage. NetApp n'a constaté aucun test de Swingbench ayant produit suffisamment d'E/S pour être une charge significative sur n'importe quelle baie AFF. Dans certains cas limités, le test OET (Order Entry Test) peut être utilisé pour évaluer le stockage du point de vue de la latence. Cela peut s'avérer utile lorsqu'une base de données a une dépendance connue en termes de latence pour des requêtes particulières. Assurez-vous que l'hôte et le réseau sont correctement configurés pour atteindre les potentiels de latence d'une baie 100 % Flash.

HammerDB

HammerDB est un outil de test de base de données qui simule les bancs d'essai TPC-C et TPC-H, entre autres. La construction d'un jeu de données suffisamment volumineux pour exécuter correctement un test peut prendre beaucoup de temps, mais elle peut constituer un outil efficace pour évaluer les performances des applications OLTP et d'entrepôt de données.

Orion

L'outil Oracle Orion a été couramment utilisé avec Oracle 9, mais il n'a pas été maintenu pour assurer la compatibilité avec les modifications apportées aux différents systèmes d'exploitation hôtes. Il est rarement utilisé avec Oracle 10 ou Oracle 11 en raison d'incompatibilités avec le système d'exploitation et la configuration du stockage.

Oracle a réécrit l'outil, qui est installé par défaut dans Oracle 12c. Bien que ce produit ait été amélioré et utilise la plupart des appels qu'une véritable base de données Oracle utilise, il n'utilise pas exactement le même chemin de code ou le même comportement d'E/S que celui utilisé par Oracle. Par exemple, la plupart des E/S Oracle sont exécutées de manière synchrone, ce qui signifie que la base de données s'arrête jusqu'à ce que les E/S soient terminées lorsque l'opération d'E/S se termine au premier plan. Le simple fait d'inonder un système de stockage d'E/S aléatoires n'est pas une reproduction de véritables E/S Oracle et n'offre pas de méthode directe pour comparer les baies de stockage ou mesurer l'impact des modifications de configuration.

Cela étant, Orion est souvent associé à des cas d'usage, comme l'évaluation générale des performances maximales d'une configuration de stockage hôte-réseau ou encore l'évaluation de l'état d'un système de stockage. Grâce à des tests rigoureux, nous pouvons concevoir des tests Orion exploitables afin de comparer les baies de stockage ou d'évaluer l'effet d'une modification de la configuration, dans la mesure où les paramètres tiennent compte des IOPS, du débit et de la latence, et tenter de répliquer fidèlement une charge de travail réaliste.

Les verrous NFSv3 obsolètes

Si un serveur de base de données Oracle tombe en panne, des verrous NFS obsolètes peuvent se présenter au redémarrage. Ce problème peut être évité en portant une attention particulière à la configuration de la résolution de nom sur le serveur.

Ce problème survient parce que la création d'un verrou et l'effacement d'un verrou utilisent deux méthodes légèrement différentes de résolution de nom. Deux processus sont impliqués, Network Lock Manager (NLM) et le client NFS. Le NLM utilise `uname n` pour déterminer le nom d'hôte, pendant que le système `rpc.statd` utilisations des processus `gethostbyname()`. Ces noms d'hôte doivent correspondre pour que le système d'exploitation efface correctement les verrous obsolètes. Par exemple, l'hôte peut rechercher des verrous

appartenant à `dbserver5`, mais les verrous ont été enregistrés par l'hôte comme `dbserver5.mydomain.org`. Si `gethostbyname()` ne renvoie pas la même valeur que `uname -a`, le processus de déverrouillage n'a pas réussi.

L'exemple de script suivant vérifie si la résolution des noms est parfaitement cohérente :

```
#! /usr/bin/perl
$uname=`uname -n`;
chomp($uname);
($name, $aliases, $addrtype, $length, @addrs) = gethostbyname $uname;
print "uname -n yields: $uname\n";
print "gethostbyname yields: $name\n";
```

Si `gethostbyname` ne correspond pas `uname`, des verrous obsolètes sont probables. Par exemple, ce résultat indique un problème potentiel :

```
uname -n yields: dbserver5
gethostbyname yields: dbserver5.mydomain.org
```

La solution est généralement trouvée en modifiant l'ordre dans lequel les hôtes apparaissent dans `/etc/hosts`. Par exemple, supposons que le fichier hosts inclut l'entrée suivante :

```
10.156.110.201  dbserver5.mydomain.org dbserver5 loghost
```

Pour résoudre ce problème, modifiez l'ordre dans lequel le nom de domaine complet et le nom d'hôte court apparaissent :

```
10.156.110.201  dbserver5 dbserver5.mydomain.org loghost
```

`gethostbyname()` renvoie maintenant le court `dbserver5` nom d'hôte, qui correspond à la sortie de `uname`. Les verrous sont donc effacés automatiquement après une panne de serveur.

Vérification de l'alignement WAFL

Un alignement WAFL correct est essentiel pour de bonnes performances. Même si ONTAP gère des blocs dans des unités de 4 Ko, ONTAP ne réalise pas forcément toutes les opérations dans des unités de 4 Ko. ONTAP prend en charge les opérations en mode bloc de différentes tailles, mais la comptabilité sous-jacente est gérée par WAFL en unités de 4 Ko.

Le terme « alignement » fait référence à la manière dont les E/S Oracle correspondent à ces unités de 4 Ko. Pour optimiser les performances, un bloc Oracle de 8 Ko doit résider sur deux blocs physiques WAFL de 4 Ko sur un disque. Si un bloc est décalé de 2 Ko, ce bloc réside dans la moitié d'un bloc de 4 Ko, dans un bloc séparé complet de 4 Ko, puis dans la moitié d'un troisième bloc de 4 Ko. Cette configuration entraîne une dégradation des performances.

L'alignement n'est pas un problème avec les systèmes de fichiers NAS. Les fichiers de données Oracle sont alignés sur le début du fichier en fonction de la taille du bloc Oracle. Par conséquent, les tailles de bloc de 8 Ko, 16 Ko et 32 Ko sont toujours alignées. Toutes les opérations de bloc sont décalées par rapport au début du fichier en unités de 4 kilo-octets.

Les LUN, en revanche, contiennent généralement au départ un type d'en-tête de pilote ou de métadonnées de système de fichiers qui crée un décalage. L'alignement est rarement un problème dans les systèmes d'exploitation modernes, car ces systèmes d'exploitation sont conçus pour des disques physiques pouvant utiliser un secteur natif de 4 Ko. De plus, ils requièrent l'alignement des E/S sur les limites de 4 Ko pour des performances optimales.

Il y a toutefois quelques exceptions. Une base de données a peut-être été migrée à partir d'un système d'exploitation plus ancien qui n'a pas été optimisé pour les E/S de 4 Ko, ou une erreur de l'utilisateur lors de la création de la partition a pu entraîner un décalage qui ne se situe pas dans des unités de 4 Ko.

Les exemples suivants sont spécifiques à Linux, mais la procédure peut être adaptée à n'importe quel système d'exploitation.

Aligné

L'exemple suivant montre une vérification d'alignement sur une seule LUN avec une seule partition.

Tout d'abord, créez la partition qui utilise toutes les partitions disponibles sur le lecteur.

```
[root@host0 iscsi]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
disklabel
Building a new DOS disklabel with disk identifier 0xb97f94c1.
Changes will remain in memory only, until you decide to write them.
After that, of course, the previous content won't be recoverable.
The device presents a logical sector size that is smaller than
the physical sector size. Aligning to a physical sector (or optimal
I/O) size boundary is recommended, or performance may be impacted.
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-10240, default 1):
Using default value 1
Last cylinder, +cylinders or +size{K,M,G} (1-10240, default 10240):
Using default value 10240
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
[root@host0 iscsi]#
```

L'alignement peut être vérifié mathématiquement à l'aide de la commande suivante :

```
[root@host0 iscsi]# fdisk -u -l /dev/sdb
Disk /dev/sdb: 10.7 GB, 10737418240 bytes
64 heads, 32 sectors/track, 10240 cylinders, total 20971520 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 4096 bytes
I/O size (minimum/optimal): 4096 bytes / 65536 bytes
Disk identifier: 0xb97f94c1
   Device Boot      Start         End      Blocks   Id  System
/dev/sdb1              32    20971519    10485744   83  Linux
```

Le résultat indique que les unités sont de 512 octets et que le début de la partition est de 32 unités. Il s'agit d'un total de 32 x 512 = 16,834 octets, soit un ensemble de blocs WAFL de 4 Ko. Cette partition est correctement alignée.

Pour vérifier que l'alignement est correct, procédez comme suit :

1. Identifier l'UUID (identifiant universel unique) de la LUN

```
FAS8040SAP::> lun show -v /vol/jfs_luns/lun0
              Vserver Name: jfs
                  LUN UUID: ed95d953-1560-4f74-9006-85b352f58fcd
                    Mapped: mapped`                 `
```

2. Entrez le shell du nœud sur le contrôleur ONTAP.

```
FAS8040SAP::> node run -node FAS8040SAP-02
Type 'exit' or 'Ctrl-D' to return to the CLI
FAS8040SAP-02> set advanced
set not found.  Type '?' for a list of commands
FAS8040SAP-02> priv set advanced
Warning: These advanced commands are potentially dangerous; use
         them only when directed to do so by NetApp
         personnel.
```

3. Démarrer les collections statistiques sur l'UUID cible identifié dans la première étape.

```
FAS8040SAP-02*> stats start lun:ed95d953-1560-4f74-9006-85b352f58fcd
Stats identifier name is 'Ind0xffffff08b9536188'
FAS8040SAP-02*>
```

4. Certaines E/S. Il est important d'utiliser le `iflag` Argument permettant de s'assurer que les E/S sont synchrones et non mises en tampon.

Faites très attention avec cette commande. Inversion du `if` et `of` les arguments détruisent les données.

```
[root@host0 iscsi]# dd if=/dev/sdb1 of=/dev/null iflag=dsync count=1000
bs=4096
1000+0 records in
1000+0 records out
4096000 bytes (4.1 MB) copied, 0.0186706 s, 219 MB/s
```

5. Arrêtez les statistiques et affichez l'histogramme d'alignement. Toutes les E/S doivent se trouver dans le `.0` Bucket, qui indique les E/S alignées sur les limites d'un bloc de 4 Ko.

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff08b9536188
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:186%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
```

Mauvais alignement

L'exemple suivant illustre un mauvais alignement des E/S :

1. Créez une partition qui ne s'aligne pas sur une limite de 4 Ko. Il ne s'agit pas d'un comportement par défaut sur les systèmes d'exploitation modernes.

```
[root@host0 iscsi]# fdisk -u /dev/sdb
Command (m for help): n
Command action
   e   extended
   p   primary partition (1-4)
p
Partition number (1-4): 1
First sector (32-20971519, default 32): 33
Last sector, +sectors or +size{K,M,G} (33-20971519, default 20971519):
Using default value 20971519
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
```

2. La partition a été créée avec un décalage de 33 secteurs au lieu du décalage de 32 par défaut. Répétez la procédure décrite à la section "Aligné". L'histogramme s'affiche comme suit :

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-
4f74-9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.0:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.1:136%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.3:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.4:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.5:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.6:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_align_histo.7:0%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:read_partial_blocks:31%
```

Le mauvais alignement est clair. Les E/S tombent principalement dans le* *`.1` godet, qui correspond au décalage attendu. Lorsque la partition a été créée, elle a été déplacée de 512 octets plus loin dans le périphérique que la valeur par défaut optimisée, ce qui signifie que l'histogramme est décalé de 512 octets.

De plus, le `read_partial_blocks` Ces statistiques ne sont pas égales à zéro, ce qui signifie que des E/S n'ont pas rempli un bloc de 4 Ko entier.

Fichiers de reprise

Les procédures décrites ici s'appliquent aux fichiers de données. Les journaux de reprise et d'archivage Oracle ont différents modèles d'E/S. Par exemple, la journalisation de reprise est un remplacement circulaire d'un seul fichier. Si la taille de bloc par défaut de 512 octets est utilisée, les statistiques d'écriture se ressemblent à ceci :

```
FAS8040SAP-02*> stats stop
StatisticsID: Ind0xffffff0468242e78
lun:ed95d953-1560-4f74-9006-85b352f58fcd:instance_uuid:ed95d953-1560-4f74-
9006-85b352f58fcd
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.0:12%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.1:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.2:4%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.3:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.4:13%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.5:6%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.6:8%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_align_histo.7:10%
lun:ed95d953-1560-4f74-9006-85b352f58fcd:write_partial_blocks:85%
```

Les E/S sont réparties dans tous les compartiments de l'histogramme, mais cela n'est pas un problème de performances. Toutefois, des taux de journalisation de reprise extrêmement élevés peuvent bénéficier d'une taille de bloc de 4 Ko. Dans ce cas, il est conseillé de vérifier que les LUN de journalisation de reprise sont correctement alignées. Cependant, cette condition n'est pas aussi importante pour de bonnes performances que l'alignement des fichiers de données.

PostgreSQL

Présentation

PostgreSQL est fourni avec des variantes incluant PostgreSQL, PostgreSQL plus et EDB Postgres Advanced Server (EPAS). PostgreSQL est généralement déployé en tant que base de données interne pour les applications multiniveaux. Il est pris en charge par les logiciels middleware courants (tels que PHP, Java, Python, Tcl/TK, ODBC, Et JDBC) et a toujours été un choix populaire pour les systèmes de gestion de bases de données open source. ONTAP constitue un excellent choix pour l'exécution des bases de données PostgreSQL et ses fonctionnalités de gestion des données fiables, performantes et efficaces.

> Cette documentation sur ONTAP et la base de données PostgreSQL remplace la base de données *TR-4770: PostgreSQL sur les meilleures pratiques ONTAP.*

Avec la croissance exponentielle des données, la gestion des données devient de plus en plus complexe pour les entreprises. Cette complexité augmente les coûts de licence, d'exploitation, de support et de maintenance. Pour réduire le coût total de possession, envisagez de passer de bases de données commerciales à des bases de données open source grâce à un stockage interne fiable et haute performance.

ONTAP est une plateforme idéale, car ONTAP est littéralement conçu pour les bases de données. De nombreuses fonctionnalités, telles que l'optimisation de la latence d'E/S aléatoire, pour la qualité de service (QoS) avancée et les fonctionnalités FlexClone de base, ont été spécialement conçues pour répondre aux besoins des charges de travail des bases de données.

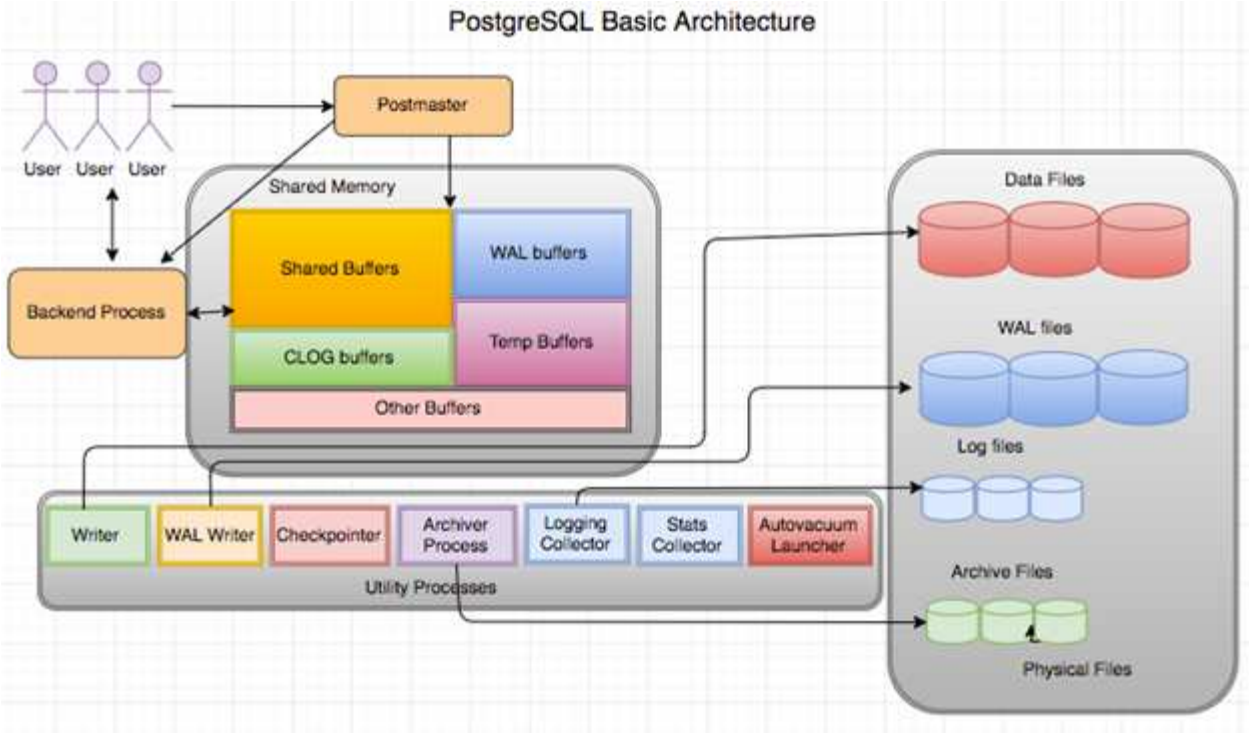
Des fonctionnalités supplémentaires, telles que les mises à niveau sans interruption (y compris le remplacement du stockage), assurent la disponibilité de vos bases de données stratégiques. Vous pouvez également bénéficier d'une reprise après incident instantanée pour les environnements volumineux via MetroCluster ou sélectionner des bases de données à l'aide de la synchronisation active SnapMirror.

Plus important encore, ONTAP offre des performances inégalées avec la possibilité de dimensionner la solution en fonction de vos besoins spécifiques. Nos systèmes haut de gamme peuvent fournir plus de 1 million d'IOPS à des latences mesurées en microsecondes. Toutefois, si vous n'avez besoin que de 100 000 IOPS, vous pouvez ajuster la taille de votre solution de stockage à l'aide d'un contrôleur plus petit, qui exécute toujours le même système d'exploitation du stockage.

Configuration de la base de données

Architecture

PostgreSQL est un SGBDR basé sur l'architecture client et serveur. Une instance PostgreSQL est appelée cluster de base de données, qui est une collection de bases de données par opposition à une collection de serveurs.



Il existe trois éléments principaux dans une base de données PostgreSQL : le postmaster, le front end (client) et le back end Le client envoie des demandes au postmaster avec des informations telles que le protocole IP et la base de données à laquelle se connecter. Le postmaster authentifie la connexion et la transmet au processus d'arrière-plan pour une communication plus poussée. Le processus back-end exécute la requête et envoie les résultats directement au frontal (client).

Une instance PostgreSQL est basée sur un modèle multiprocessus au lieu d'un modèle multithread. Il génère plusieurs processus pour différents travaux, et chaque processus possède sa propre fonctionnalité. Les principaux processus incluent le processus client, le processus WAL writer, le processus Background writer et le processus checkpointer :

- Lorsqu'un processus client (premier plan) envoie des demandes de lecture ou d'écriture à l'instance PostgreSQL, il ne lit pas ou n'écrit pas les données directement sur le disque. Il met d'abord en mémoire tampon les données dans des tampons partagés et des tampons WAL (Write-Ahead Logging).
- Un processus WAL writer manipule le contenu des tampons partagés et des tampons WAL pour écrire dans les journaux WAL. Les journaux WAL sont généralement des journaux de transaction de PostgreSQL et sont écrits de manière séquentielle. Par conséquent, pour améliorer le temps de réponse de la base de données, PostgreSQL écrit d'abord dans les journaux de transactions et reconnaît le client.
- Pour mettre la base de données dans un état cohérent, le processus de l'enregistreur d'arrière-plan vérifie périodiquement la présence de pages sales dans le tampon partagé. Il purge ensuite les données sur les fichiers de données stockés sur des volumes NetApp ou des LUN.
- Le processus CheckPointer s'exécute également périodiquement (moins fréquemment que le processus d'arrière-plan) et empêche toute modification des tampons. Il signale au processus d'écriture WAL d'écrire et de vider l'enregistrement de point de contrôle à la fin des journaux WAL stockés sur le disque NetApp. Il signale également au processus d'écriture d'arrière-plan d'écrire et de vider toutes les pages sales sur le disque.

Paramètres d'initialisation

Vous créez un nouveau cluster de base de données à l'aide de `initdb` programme. An

`initdb` script crée les fichiers de données, les tables système et les bases de données modèles (template0 et template1) qui définissent le cluster.

La base de données de modèles représente une base de données de stock. Il contient des définitions pour les tables système, les vues standard, les fonctions et les types de données. `pgdata` sert d'argument à l' `initdb` script qui spécifie l'emplacement du cluster de base de données.

Tous les objets de base de données dans PostgreSQL sont gérés en interne par les OID respectives. Les tables et les index sont également gérés par des OID individuelles. Les relations entre les objets de base de données et leurs OID respectives sont stockées dans les tables de catalogue système appropriées, selon le type d'objet. Par exemple, les OID des bases de données et des tables de segment de mémoire sont stockées dans `pg_database` et `pg_class, respectivement. Vous pouvez déterminer les OID en émettant des requêtes sur le client PostgreSQL.

Chaque base de données a ses propres tables et fichiers d'index qui sont limités à 1 Go. Chaque table a deux fichiers associés, avec le suffixe respectivement `_fsm` et `_vm`. Ils sont appelés carte de l'espace libre et carte de visibilité. Ces fichiers stockent les informations relatives à la capacité d'espace libre et ont une visibilité sur chaque page du fichier de table. Les index ne disposent que de cartes d'espace libre individuelles et ne disposent pas de cartes de visibilité.

Le `pg_xlog/pg_wal` le répertoire contient les journaux d'écriture anticipée. Des journaux d'écriture anticipée sont utilisés pour améliorer la fiabilité et les performances de la base de données. Chaque fois que vous mettez à jour une ligne dans une table, PostgreSQL écrit d'abord la modification dans le journal d'écriture anticipée, puis écrit les modifications sur les pages de données réelles sur un disque. Le `pg_xlog` le répertoire contient généralement plusieurs fichiers, mais initdb ne crée que le premier. Des fichiers supplémentaires sont ajoutés si nécessaire. Chaque fichier xlog fait 16 Mo de long.

Paramètres

Il existe plusieurs configurations de réglage PostgreSQL qui peuvent améliorer les performances.

Les paramètres les plus utilisés sont les suivants :

- `max_connections = <num>`: Le nombre maximal de connexions de base de données à avoir en même temps. Utilisez ce paramètre pour limiter l'échange sur le disque et l'arrêt des performances. Selon les besoins de votre application, vous pouvez également régler ce paramètre pour les paramètres du pool de connexions.
- `shared_buffers = <num>`: La méthode la plus simple pour améliorer les performances de votre serveur de base de données. La valeur par défaut est faible pour la plupart des matériels modernes. Il est défini pendant le déploiement à environ 25 % de la RAM disponible sur le système. Ce paramètre varie en fonction de la façon dont il fonctionne avec des instances de base de données particulières ; vous devrez peut-être augmenter ou diminuer les valeurs par tâtonnement et erreur. Cependant, le réglage haut risque de dégrader les performances.
- `effective_cache_size = <num>`: Cette valeur indique à l'optimiseur de PostgreSQL la quantité de mémoire disponible pour la mise en cache des données et aide à déterminer si un index doit être utilisé. Une valeur plus élevée augmente la probabilité d'utiliser un index. Ce paramètre doit être défini sur la quantité de mémoire allouée à `shared_buffers` Plus la quantité de cache du système d'exploitation disponible. Cette valeur représente souvent plus de 50 % de la mémoire système totale.
- `work_mem = <num>`: Ce paramètre contrôle la quantité de mémoire à utiliser dans les opérations de tri et les tables de hachage. Si vous effectuez un tri important dans votre application, vous devrez peut-être augmenter la quantité de mémoire, mais soyez prudent. Ce n'est pas un paramètre à l'échelle du système,

mais un paramètre par opération. Si une requête complexe comporte plusieurs opérations de tri, elle utilise plusieurs unités de mémoire Work_mem et plusieurs back end peuvent le faire simultanément. Cette requête peut souvent amener votre serveur de base de données à échanger si la valeur est trop élevée. Cette option était auparavant appelée sort_mem dans les anciennes versions de PostgreSQL.

- `fsync = <boolean> (on or off)`: Ce paramètre détermine si toutes vos pages WAL doivent être synchronisées sur le disque à l'aide de fsync() avant qu'une transaction ne soit validée. Sa désactivation peut parfois améliorer les performances d'écriture et son activation renforce la protection contre le risque de corruption en cas de panne du système.
- `checkpoint_timeout`: Le processus de point de contrôle vide les données validées sur le disque. Cela implique de nombreuses opérations de lecture/écriture sur le disque. La valeur est définie en secondes et les valeurs inférieures réduisent le temps de reprise après incident et l'augmentation des valeurs peut réduire la charge sur les ressources système en réduisant les appels au point de contrôle. En fonction de la criticité de l'application, de l'utilisation et de la disponibilité de la base de données, définissez la valeur de Checkpoint_timeout.
- `commit_delay = <num>` et `commit_siblings = <num>`: Ces options sont utilisées ensemble pour aider à améliorer les performances en écrivant plusieurs transactions qui sont exécutées simultanément. Si plusieurs objets commit_frames sont actifs à l'instant où votre transaction est validée, le serveur attend les microsecondes commit_delay pour essayer de valider plusieurs transactions à la fois.
- `max_worker_processes / max_parallel_workers`: Configurer le nombre optimal de travailleurs pour les processus. Max_Parallel_workers correspond au nombre de CPU disponibles. Selon la conception de l'application, les requêtes peuvent nécessiter un nombre réduit de collaborateurs pour les opérations parallèles. Il est préférable de conserver la même valeur pour les deux paramètres, mais d'ajuster la valeur après le test.
- `random_page_cost = <num>`: Cette valeur contrôle la façon dont PostgreSQL affiche les lectures de disque non séquentielles. Une valeur plus élevée signifie que PostgreSQL est plus susceptible d'utiliser une analyse séquentielle au lieu d'une analyse d'index, indiquant que votre serveur a des disques rapides Modifier ce paramètre après avoir évalué d'autres options telles que l'optimisation basée sur un plan, l'aspiration, l'indexation pour modifier les requêtes ou le schéma.
- `effective_io_concurrency = <num>`: Ce paramètre définit le nombre d'opérations d'E/S de disque simultanées que PostgreSQL tente d'exécuter simultanément. L'augmentation de cette valeur augmente le nombre d'opérations d'E/S que toute session PostgreSQL individuelle tente d'initier en parallèle. La plage autorisée est comprise entre 1 et 1,000, ou zéro pour désactiver l'émission de demandes d'E/S asynchrones. Actuellement, ce paramètre n'affecte que les analyses de tas bitmap. Les disques SSD et les autres systèmes de stockage basés sur la mémoire (NVMe) peuvent souvent traiter un grand nombre de requêtes simultanées. Le meilleur choix peut donc se situer dans les centaines.

Consultez la documentation PostgreSQL pour obtenir une liste complète des paramètres de configuration PostgreSQL.

TOASTS

TOAST est l'acronyme de Oversized-Attribute Storage technique. PostgreSQL utilise une taille de page fixe (généralement 8 Ko) et ne permet pas aux blocs de données de couvrir plusieurs pages. Par conséquent, il n'est pas possible de stocker directement des valeurs de champ importantes. Lorsque vous essayez de stocker une ligne qui dépasse cette taille, TOAST divise les données de grandes colonnes en « morceaux » plus petits et les stocke dans une table de TOASTS.

Les grandes valeurs des attributs toastés sont extraites (si elles sont sélectionnées) uniquement au moment où le jeu de résultats est envoyé au client. La table elle-même est beaucoup plus petite et peut contenir plus de lignes dans le cache du tampon partagé qu'elle ne le pouvait sans stockage hors ligne (TOAST).

VIDE

En mode PostgreSQL normal, les blocs de données supprimés ou rendus obsolètes par une mise à jour ne sont pas physiquement supprimés de leur table ; ils restent présents jusqu'à ce que LE VIDE soit exécuté. Par conséquent, vous devez faire fonctionner le VIDE régulièrement, en particulier sur les tables fréquemment mises à jour. L'espace qu'il occupe doit ensuite être récupéré pour réutilisation par de nouvelles lignes, afin d'éviter une panne d'espace disque. Cependant, il ne renvoie pas l'espace vers le système d'exploitation.

L'espace libre dans une page n'est pas fragmenté. VIDE réécrit le bloc entier, en empaquant efficacement les lignes restantes et en laissant un seul bloc contigu d'espace libre dans une page.

En revanche, LE VIDE COMPLET composera activement les tables en écrivant une version complètement nouvelle du fichier table sans espace mort. Cette action réduit la taille de la table mais peut prendre un certain temps. Elle nécessite également de l'espace disque supplémentaire pour la nouvelle copie de la table jusqu'à ce que l'opération soit terminée. L'objectif du VIDE DE routine est d'éviter toute activité de VIDE COMPLET. Ce processus permet non seulement de conserver les tables à leur taille minimale, mais également de conserver une utilisation régulière de l'espace disque.

Espaces de stockage

Deux tablespaces sont créés automatiquement lorsque le cluster de base de données est initialisé.

Le `pg_global` l'espace table est utilisé pour les catalogues système partagés. Le `pg_default` tablespace est l'espace table par défaut des bases de données templage1 et template0. Si la partition ou le volume sur lequel le cluster a été initialisé est à court d'espace et ne peut pas être étendu, un espace table peut être créé sur une partition différente et utilisé jusqu'à ce que le système puisse être reconfiguré.

Un index très utilisé peut être placé sur un disque rapide et hautement disponible, comme un périphérique SSD. Par ailleurs, une table qui stocke des données archivées rarement utilisées ou non critiques pour les performances peut être stockée sur un système sur disque moins onéreux et plus lent, tel que des disques SAS ou SATA.

Les tablespaces font partie du cluster de base de données et ne peuvent pas être traités comme un ensemble autonome de fichiers de données. Elles dépendent des métadonnées contenues dans le répertoire de données principal et ne peuvent donc pas être reliées à un autre cluster de base de données ou sauvegardées individuellement. De même, si vous perdez un espace de table (suite à la suppression d'un fichier, à une panne de disque, etc.), le cluster de base de données peut devenir illisible ou ne pas démarrer. Le fait de placer un tablespace sur un système de fichiers temporaire, tel qu'un disque RAM, risque de nuire à la fiabilité de l'ensemble du cluster.

Une fois créé, un espace table peut être utilisé à partir de n'importe quelle base de données si l'utilisateur demandeur dispose de privilèges suffisants. PostgreSQL utilise des liens symboliques pour simplifier l'implémentation des tablespaces. PostgreSQL ajoute une ligne au `pg_tablespace` Tableau (table à l'échelle du cluster) et attribue un nouvel identifiant d'objet (OID) à cette ligne. Enfin, le serveur utilise l'OID pour créer un lien symbolique entre votre cluster et le répertoire donné. Le répertoire `$PGDATA/pg_tblspc` contient des liens symboliques pointant vers chacun des tablespaces non intégrés définis dans le cluster.

Configuration de stockage sous-jacente

NFS

Les bases de données PostgreSQL peuvent être hébergées sur les systèmes de fichiers

NFSv3 ou NFSv4. La meilleure option dépend de facteurs extérieurs à la base de données.

Par exemple, le comportement de verrouillage NFSv4 peut être préférable dans certains environnements en cluster. (Voir "ici" pour plus d'informations)

Dans le cas contraire, les fonctionnalités de la base de données doivent être proches des mêmes, y compris les performances. La seule exigence est l'utilisation du `hard` option de montage. Ceci est nécessaire pour garantir que les délais d'expiration ne produisent pas d'erreurs d'E/S irrécupérables.

Si NFSv4 est choisi en tant que protocole, NetApp recommande d'utiliser NFSv4.1. Certaines améliorations fonctionnelles du protocole NFSv4 dans NFSv4.1 améliorent la résilience sur NFSv4.0.

Utilisez les options de montage suivantes pour les charges de travail de base de données générales :

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=65536,wsize=65536
```

Si des E/S séquentielles lourdes sont attendues, la taille du transfert NFS peut être augmentée comme décrit dans la section suivante.

Tailles de transfert NFS

Par défaut, ONTAP limite la taille des E/S NFS à 64 Ko.

Les E/S aléatoires utilisent la plupart des applications et bases de données une taille de bloc bien inférieure à la taille maximale de 64 Ko. Les E/S de blocs volumineux sont généralement parallélisées de sorte que le maximum de 64 Ko ne limite pas non plus l'obtention d'une bande passante maximale.

Dans certains cas, le maximum de 64 000 charges de travail entraîne une limitation. En particulier, les opérations à thread unique, telles que les opérations de sauvegarde ou de restauration, ou encore les analyses de table complète de base de données s'exécutent plus rapidement et plus efficacement si la base de données peut exécuter moins d'E/S, mais plus volumineuses. La taille optimale de gestion des E/S pour ONTAP est de 256 Ko.

La taille maximale de transfert pour un SVM ONTAP donné peut être modifiée comme suit :

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

> Ne réduisez jamais la taille de transfert maximale autorisée sur ONTAP en dessous de la valeur de rsize/wsize des systèmes de fichiers NFS actuellement montés. Cela peut provoquer des blocages ou même une corruption des données avec certains systèmes d'exploitation. Par exemple, si les clients NFS sont actuellement définis sur une taille rsize/wsize de 65536, la taille maximale du transfert ONTAP peut être ajustée entre 65536 et 1048576 sans effet car les clients eux-mêmes sont limités. Réduire la taille de transfert maximale en dessous de 65536 peut endommager la disponibilité ou les données.

Une fois la taille de transfert augmentée au niveau ONTAP, les options de montage suivantes sont utilisées :

```
rw,hard,nointr,bg,vers=[3|4],proto=tcp,rsize=262144,wsize=262144
```

Tables d'emplacements TCP NFSv3

Si NFSv3 est utilisé avec Linux, il est essentiel de définir correctement les tables d'emplacements TCP.

Les tables d'emplacements TCP sont l'équivalent NFSv3 de la profondeur de file d'attente de l'adaptateur de bus hôte (HBA). Ces tableaux contrôlent le nombre d'opérations NFS qui peuvent être en attente à la fois. La valeur par défaut est généralement 16, un chiffre bien trop faible pour assurer des performances optimales. Le problème inverse se produit sur les noyaux Linux plus récents : la limite de la table des emplacements TCP augmente automatiquement par envoi de demandes, jusqu'à atteindre le niveau de saturation du serveur NFS.

Pour des performances optimales et pour éviter les problèmes de performances, ajustez les paramètres du noyau qui contrôlent les tables d'emplacements TCP.

Exécutez le `sysctl -a | grep tcp.*.slot_table` et observez les paramètres suivants :

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tous les systèmes Linux doivent inclure `sunrpc.tcp_slot_table_entries`, mais seulement certains incluent `sunrpc.tcp_max_slot_table_entries`. Ils doivent tous deux être réglés sur 128.

> Si vous ne définissez pas ces paramètres, vous risquez d'avoir des effets importants sur les performances. Dans certains cas, les performances sont limitées car le système d'exploitation linux n'émet pas suffisamment d'E/S. Dans d'autres cas, les latences d'E/S augmentent à mesure que le système d'exploitation linux tente d'émettre plus d'E/S que ce qui peut être traité.

SAN

Les bases de données PostgreSQL avec SAN sont généralement hébergées sur des systèmes de fichiers xfs, mais d'autres peuvent être utilisées si elles sont prises en charge par le fournisseur du système d'exploitation

Même si un seul LUN peut généralement prendre en charge jusqu'à 100 000 IOPS, les bases de données exigeantes en E/S nécessitent généralement l'utilisation de LVM avec répartition.

Marquage LVM

Avant l'ère des disques Flash, la répartition était utilisée pour surmonter les limites de performances des disques rotatifs. Par exemple, si un système d'exploitation doit effectuer une opération de lecture de 1 Mo, la lecture de ce 1 Mo de données à partir d'un seul disque demande beaucoup de tête de lecture lorsque le transfert des 1 Mo est lent. Si ce 1 Mo de données a été réparti sur 8 LUN, le système d'exploitation pourrait exécuter huit opérations de lecture de 128 K en parallèle et réduire le temps nécessaire au transfert de 1 Mo.

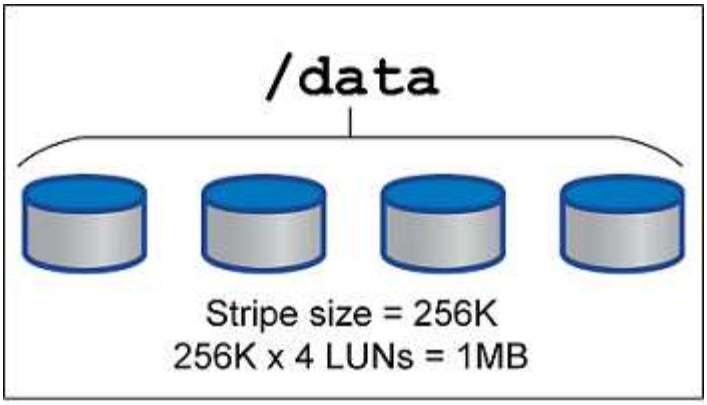
Le striping avec des disques rotatifs était plus difficile, car le modèle d'E/S devait être connu à l'avance. Si la répartition n'a pas été correctement réglée pour les véritables modèles d'E/S, les configurations à bandes risquent d'endommager les performances. Avec les bases de données Oracle, et en particulier les configurations 100 % Flash, le striping est beaucoup plus facile à configurer et a fait ses preuves pour améliorer considérablement les performances.

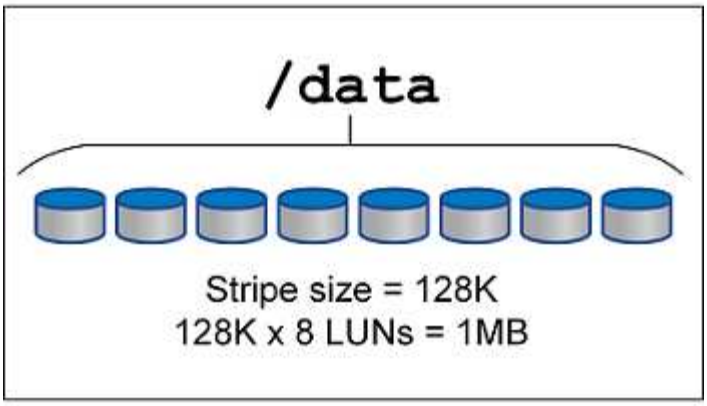
Par défaut, les gestionnaires de volumes logiques, tels que la bande Oracle ASM, ne le font pas pour le système d'exploitation natif LVM. Certaines lient plusieurs LUN ensemble en tant que périphérique concaténé. Résultat : des fichiers de données existent sur un seul périphérique LUN. Ceci provoque des points chauds. Les autres implémentations LVM prennent par défaut en charge les extensions distribuées. Cette méthode est similaire à la répartition, mais elle est plus grossière. Les LUN du groupe de volumes sont tranchées en grandes parties, appelées extensions et généralement mesurées en plusieurs mégaoctets. Ensuite, les volumes logiques sont distribués sur ces extensions. Il en résulte des E/S aléatoires sur un fichier qui doit être bien réparti entre les LUN, mais les opérations d'E/S séquentielles ne sont pas aussi efficaces qu'elles pourraient l'être.

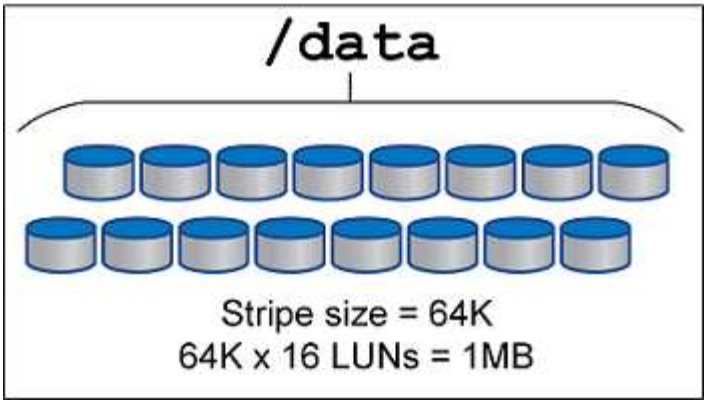
Les E/S des applications exigeantes en performances sont presque toujours de (a) en unités de taille de bloc de base ou (b) d'un mégaoctet.

L'objectif principal d'une configuration à bandes est de s'assurer que les E/S de fichier unique peuvent être exécutées comme une seule unité, et que les E/S de plusieurs blocs, d'une taille de 1 Mo, peuvent être parallélisées de façon homogène sur toutes les LUN du volume réparti. Cela signifie que la taille de bande ne doit pas être inférieure à la taille du bloc de base de données, et que la taille de bande multipliée par le nombre de LUN doit être de 1 Mo.

La figure suivante présente trois options possibles pour le réglage de la taille et de la largeur des bandes. Le nombre de LUN est sélectionné pour répondre aux exigences de performances comme décrit ci-dessus, mais dans tous les cas, le total des données dans une seule bande est de 1 Mo.







Protection des données

Protection Ddta native

L'un des principaux aspects de la conception du stockage est l'activation de la protection pour les volumes PostgreSQL. Les clients peuvent protéger leurs bases de données PostgreSQL en utilisant l'approche dump ou en utilisant des sauvegardes de système de fichiers. Cette section décrit les différentes approches de sauvegarde de bases de données individuelles ou de l'ensemble du cluster.

Il existe trois approches de sauvegarde des données PostgreSQL :

- Dump SQL Server
- Sauvegarde au niveau du système de fichiers

- Archivage continu

L'idée derrière la méthode de vidage de SQL Server est de générer un fichier avec des commandes SQL Server qui, une fois renvoyées au serveur, peuvent recréer la base de données telle qu'elle était au moment de la sauvegarde. PostgreSQL fournit les programmes utilitaires `pg_dump` et `pg_dump_all` pour la création de sauvegardes individuelles et au niveau du cluster. Ces vidages sont logiques et ne contiennent pas suffisamment d'informations pour être utilisés par la relecture WAL.

Une autre stratégie de sauvegarde consiste à utiliser une sauvegarde au niveau du système de fichiers, dans laquelle les administrateurs copient directement les fichiers utilisés par PostgreSQL pour stocker les données dans la base de données. Cette méthode s'effectue en mode hors ligne : la base de données ou le cluster doit être arrêté. Une autre alternative est d'utiliser `pg_basebackup` Pour exécuter une sauvegarde de diffusion à chaud de la base de données PostgreSQL.

Snapshots

Les sauvegardes basées sur des copies Snapshot avec PostgreSQL requièrent la configuration de snapshots pour les fichiers de données, les fichiers WAL et les fichiers WAL archivés afin d'assurer une restauration complète ou instantanée.

Pour les bases de données PostgreSQL, la durée moyenne de sauvegarde avec des copies Snapshot est comprise entre quelques secondes et quelques minutes. Cette vitesse de sauvegarde est 60 à 100 fois plus rapide que `pg_basebackup` et d'autres approches de sauvegarde basées sur le système de fichiers.

Les copies Snapshot situées sur un système de stockage NetApp peuvent être à la fois cohérentes après panne et cohérentes au niveau des applications. Un snapshot cohérent après panne est créé sur un système de stockage sans interrompre la base de données, tandis qu'un Snapshot cohérent avec les applications est créé lorsque la base de données est en mode de sauvegarde. NetApp garantit également que les copies Snapshot suivantes sont des sauvegardes incrémentielles à l'infini pour promouvoir les économies de stockage et l'efficacité réseau.

Comme les snapshots sont rapides et n'affectent pas les performances du système, vous pouvez planifier plusieurs copies Snapshot chaque jour au lieu de créer une sauvegarde quotidienne comme avec les autres technologies de sauvegarde en streaming. Lorsqu'une opération de restauration et de restauration est nécessaire, le temps d'interruption du système est réduit grâce à deux fonctionnalités clés :

- Avec la technologie de restauration des données NetApp SnapRestore, la restauration s'exécute en quelques secondes.
- Les objectifs de point de restauration (RPO) agressifs signifient qu'il faut moins de journaux de base de données et que la restauration par progression est également accélérée.

Pour sauvegarder PostgreSQL, vous devez vous assurer que les volumes de données sont protégés simultanément avec WAL (groupe de cohérence) et les journaux archivés. Lorsque vous utilisez la technologie Snapshot pour copier des fichiers WAL, assurez-vous de les exécuter `pg_stop` Pour vider toutes les entrées WAL qui doivent être archivées. Si vous videz les entrées WAL pendant la restauration, il vous suffit d'arrêter la base de données, de démonter ou de supprimer le répertoire de données existant et d'effectuer une opération SnapRestore sur le stockage. Une fois la restauration terminée, vous pouvez monter le système et le ramener à son état actuel. Pour la restauration instantanée, vous pouvez également restaurer les journaux WAL et d'archivage ; PostgreSQL décide alors du point le plus cohérent et le récupère automatiquement.

Les groupes de cohérence sont une fonctionnalité de ONTAP recommandée lorsque plusieurs volumes sont montés sur une seule instance ou une base de données avec plusieurs tablespaces. Une copie Snapshot de groupe de cohérence garantit que tous les volumes sont regroupés et protégés. Pour gérer efficacement un

groupe de cohérence, ONTAP vous pouvez même le cloner et créer une copie d'instance d'une base de données à des fins de test ou de développement.

Logiciels de protection des données

Le plug-in NetApp SnapCenter pour les bases de données PostgreSQL, associé aux technologies Snapshot et NetApp FlexClone, vous offre les avantages suivants :

- Sauvegarde et restauration rapides.
- Clones compacts.
- La possibilité de mettre en place un système de reprise d'activité rapide et efficace.

> Vous pouvez choisir les partenaires de sauvegarde premium de NetApp, tels que Veeam Software et CommVault dans les circonstances suivantes :
>
> - Gestion des workloads dans un environnement hétérogène
> - Stocker les sauvegardes dans le cloud ou sur bande pour les conserver à long terme
> - Prise en charge d'un large éventail de versions et de types de systèmes d'exploitation

Le plug-in SnapCenter pour PostgreSQL est un plug-in pris en charge par la communauté et la configuration et la documentation sont disponibles sur le magasin d'automatisation NetApp. Grâce à SnapCenter, l'utilisateur peut sauvegarder la base de données, cloner et restaurer les données à distance.

VMware

VMware vSphere avec ONTAP

VMware vSphere avec ONTAP

ONTAP a servi de solution de stockage de premier plan pour VMware vSphere et, plus récemment, pour les environnements Cloud Foundation depuis son introduction dans le data Center moderne en 2002. Elle continue d'introduire des fonctionnalités innovantes qui simplifient la gestion et réduisent les coûts.

Ce document présente la solution ONTAP pour vSphere et met en avant les dernières informations sur les produits et les meilleures pratiques pour rationaliser le déploiement, limiter les risques et simplifier la gestion.

> Cette documentation remplace les rapports techniques *TR-4597 : VMware vSphere pour ONTAP*

Les meilleures pratiques complètent d'autres documents, tels que des guides et des listes de compatibilité. Ils sont développés en fonction de tests effectués en laboratoire et d'une vaste expérience sur le terrain par les ingénieurs et les clients NetApp. Non seulement elles sont les seules pratiques prises en charge dans chaque environnement, mais elles constituent généralement les solutions les plus simples qui répondent aux besoins de la plupart des clients.

Ce document est axé sur les fonctionnalités des dernières versions d'ONTAP (9.x) exécutées sur vSphere 7.0 ou version ultérieure. Consultez "Matrice d'interopérabilité (IMT)" et "Guide de compatibilité VMware" pour plus d'informations sur des versions spécifiques.

Pourquoi choisir ONTAP pour VMware vSphere ?

Les clients choisissent en toute confiance ONTAP pour vSphere pour les solutions de stockage SAN et NAS. La nouvelle architecture de stockage désagrégée simplifiée, présente dans les dernières baies All SAN, offre une expérience simplifiée familière aux administrateurs de stockage SAN tout en conservant la plupart des intégrations et des fonctionnalités des systèmes ONTAP traditionnels. Les systèmes ONTAP offrent une protection exceptionnelle des snapshots et des outils de gestion robustes. En déchargeant les fonctions vers un stockage dédié, ONTAP maximise les ressources de l'hôte, réduit les coûts et maintient des performances optimales. De plus, les charges de travail peuvent être facilement migrées à l'aide de Storage vMotion sur VMFS, NFS ou vVols.

Avantages de l'utilisation de ONTAP pour vSphere

De nombreuses raisons ont poussé des dizaines de milliers de clients à choisir ONTAP comme solution de stockage pour vSphere, par exemple un système de stockage unifié prenant en charge les protocoles SAN et NAS, des fonctionnalités robustes de protection des données à l'aide de copies Snapshot compactes et une multitude d'outils pour vous aider à gérer les données applicatives. En utilisant un système de stockage distinct de l'hyperviseur, vous pouvez décharger de nombreuses fonctions et optimiser votre investissement dans les systèmes hôtes vSphere. En plus de s'assurer que les ressources de vos hôtes sont concentrées sur les charges de travail applicatives, vous évitez également l'impact aléatoire sur les performances des

applications en provenance des opérations de stockage.

L'utilisation ONTAP avec vSphere est une excellente combinaison qui vous permet de réduire les dépenses liées au matériel hôte et aux logiciels VMware. Vous pouvez également protéger vos données à moindre coût avec des performances élevées et constantes. Étant donné que les charges de travail virtualisées sont mobiles, vous pouvez explorer différentes approches à l'aide de Storage vMotion pour déplacer des machines virtuelles entre des banques de données VMFS, NFS ou vVols , le tout sur le même système de stockage.

Voici les facteurs clés que les clients apprécient aujourd'hui :

- **Stockage unifié.** Les systèmes exécutant ONTAP sont unifiés de plusieurs manières importantes. À l'origine, cette approche faisait référence aux protocoles NAS et SAN, et ONTAP continue d'être une plate-forme de premier plan pour le SAN, avec sa force initiale dans le NAS. Dans le monde vSphere, cette approche pourrait également signifier un système unifié pour l'infrastructure de bureau virtuel (VDI) ainsi que pour l'infrastructure de serveur virtuel (VSI). Les systèmes exécutant ONTAP sont généralement moins chers pour VSI que les baies d'entreprise traditionnelles et disposent pourtant de capacités d'efficacité de stockage avancées pour gérer VDI dans le même système. ONTAP unifie également une variété de supports de stockage, des SSD aux SATA, et peut facilement les étendre au cloud. Il n'est pas nécessaire d'acheter un système d'exploitation de stockage pour les performances, un autre pour les archives et encore un autre pour le cloud. ONTAP les relie tous ensemble.
- **Baie SAN (ASA).** Les derniers systèmes ONTAP ASA (à partir des modèles A1K, A90, A70, A50, A30 et A20) reposent sur une nouvelle architecture de stockage qui élimine le modèle de stockage ONTAP classique utilisé pour la gestion des agrégats et des volumes. Comme il n'existe aucun partage de système de fichiers, les volumes ne sont pas nécessaires. Tout le stockage rattaché à une paire HA est traité comme une zone de disponibilité du stockage commune (SAZ) dans laquelle les LUN et les espaces de noms NVMe sont provisionnés en tant que « unités de stockage » (MU). Les derniers systèmes ASA sont conçus pour être faciles à gérer et offrent une expérience familière aux administrateurs de stockage SAN. Cette nouvelle architecture est idéale pour les environnements vSphere, car elle facilite la gestion des ressources de stockage et simplifie l'expérience des administrateurs de stockage SAN. L'architecture ASA prend également en charge la dernière technologie NVMe over Fabrics (NVMe-of) qui améliore encore les performances et l'évolutivité des workloads vSphere.
- **Technologie Snapshot.** ONTAP a été le premier à proposer une technologie Snapshot pour la protection des données et reste la plus avancée du secteur. Cette approche peu gourmande en espace pour la protection des données a été étendue pour prendre en charge les API VMware vSphere pour l'intégration de baies (VAAI). Cette intégration vous permet de tirer parti des fonctionnalités Snapshot de ONTAP pour les opérations de sauvegarde et de restauration, réduisant ainsi l'impact sur votre environnement de production. Cette approche vous permet également d'utiliser des snapshots pour une restauration rapide des machines virtuelles, réduisant ainsi le temps et les efforts nécessaires à la restauration des données. De plus, la technologie Snapshot de ONTAP est intégrée aux solutions VLSR (Live site Recovery Manager) de VMware, offrant ainsi une stratégie complète de protection des données pour votre environnement virtualisé.
- **Gestion basée sur des politiques de volumes virtuels et de stockage.** NetApp a été l'un des premiers partenaires de conception de VMware dans le développement de vSphere Virtual Volumes (vVols), fournissant des contributions architecturales et un support précoce pour vVols et VMware vSphere APIs for Storage Awareness (VASA). Non seulement cette approche a apporté une gestion granulaire du stockage des machines virtuelles à VMFS, mais elle a également pris en charge l'automatisation du provisionnement du stockage via une gestion basée sur des politiques de stockage. Cette approche permet aux architectes de stockage de concevoir des pools de stockage dotés de différentes capacités qui peuvent être facilement utilisées par les administrateurs de machines virtuelles. ONTAP est le leader du secteur du stockage en termes d'échelle vVol, prenant en charge des centaines de milliers de vVols dans un seul cluster, tandis que les fournisseurs de baies d'entreprise et de baies flash plus petites prennent en charge quelques milliers de vVols par baie. NetApp est également à l'origine de l'évolution de la gestion granulaire des machines virtuelles avec des fonctionnalités à venir.

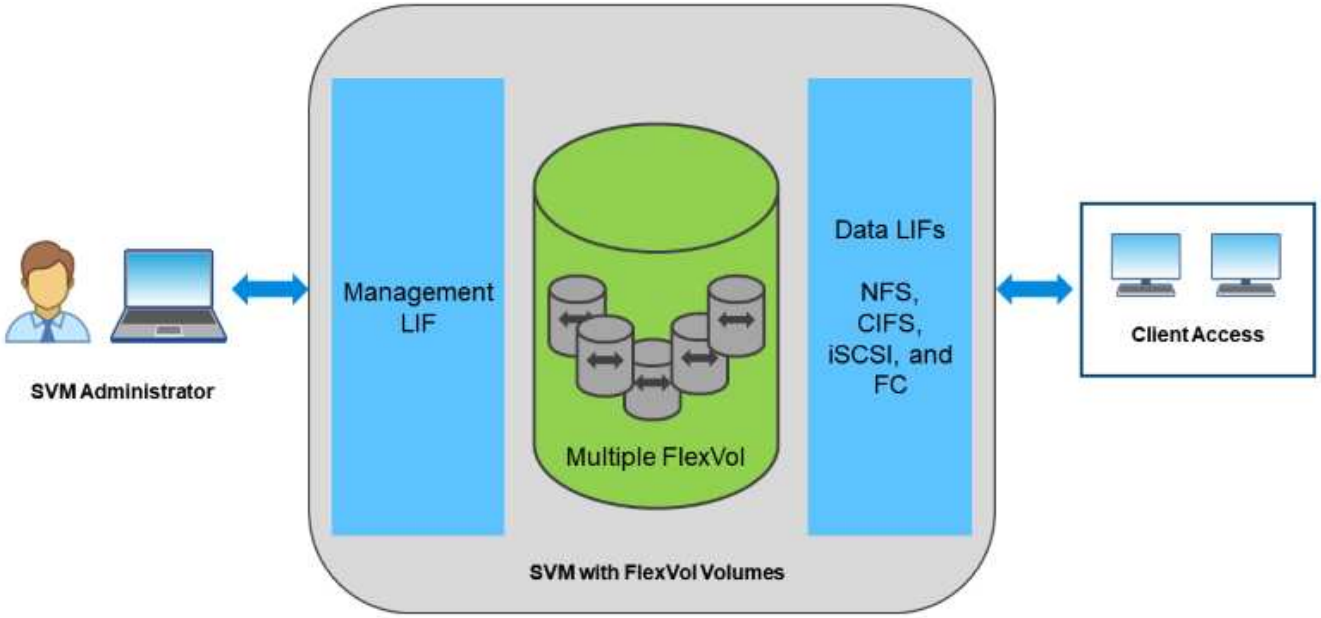
- **Efficacité de stockage.** Bien que NetApp ait été le premier à proposer la déduplication pour les charges de travail de production, cette innovation n'était ni la première ni la dernière dans ce domaine. Tout a commencé avec des instantanés, un mécanisme de protection des données peu encombrant sans effet sur les performances, ainsi que la technologie FlexClone pour créer instantanément des copies en lecture/écriture de machines virtuelles à des fins de production et de sauvegarde. NetApp a ensuite proposé des fonctionnalités en ligne, notamment la déduplication, la compression et la déduplication zéro bloc, pour tirer le meilleur parti du stockage des SSD coûteux. ONTAP a également ajouté la possibilité de regrouper des opérations d'E/S et des fichiers plus petits dans un bloc de disque à l'aide de la compaction. La combinaison de ces capacités a permis aux clients de réaliser des économies allant jusqu'à 5:1 pour VSI et jusqu'à 30:1 pour VDI. La dernière génération de systèmes ONTAP inclut également la compression et la déduplication accélérées par le matériel, ce qui peut encore améliorer l'efficacité du stockage et réduire les coûts. Cette approche vous permet de stocker plus de données dans moins d'espace, réduisant ainsi le coût global du stockage et améliorant les performances. NetApp est tellement confiant dans ses capacités d'efficacité de stockage qu'il propose un lien : https://www.netapp.com/pdf.html?item=/media/79014-ng-937-Efficiency-Guarantee-Customer-Flyer.pdf [Garantie d'efficacité^].
- **Multilocation.** ONTAP est depuis longtemps un leader en matière de multilocation, vous permettant de créer plusieurs machines virtuelles de stockage (SVM) sur un seul cluster. Cette approche vous permet d'isoler les charges de travail et de fournir différents niveaux de service à différents locataires, ce qui la rend idéale pour les fournisseurs de services et les grandes entreprises. La dernière génération de systèmes ONTAP inclut également la prise en charge de la gestion de la capacité des locataires. Cette fonctionnalité vous permet de définir des limites de capacité pour chaque locataire, garantissant qu'aucun locataire ne peut consommer toutes les ressources disponibles. Cette approche permet de garantir que tous les locataires reçoivent le niveau de service qu'ils attendent, tout en offrant un niveau élevé de sécurité et d'isolement entre les locataires. De plus, les capacités multi-tenant d'ONTAP sont intégrées à la plate-forme vSphere de VMware, vous permettant de gérer et de surveiller facilement votre environnement virtualisé via "Les outils ONTAP pour VMware vSphere" et "Informations exploitables sur l'infrastructure de données" .
- **Cloud hybride.** Qu'elles soient utilisées pour un cloud privé sur site, une infrastructure de cloud public ou un cloud hybride combinant le meilleur des deux, les solutions ONTAP vous aident à créer votre structure de données pour rationaliser et optimiser la gestion des données. Commencez par des systèmes 100 % flash hautes performances, puis associez-les à des systèmes de stockage sur disque ou dans le cloud pour la protection des données et le calcul dans le cloud. Choisissez parmi Azure, AWS, IBM ou Google Cloud pour optimiser vos coûts et éviter le blocage. Bénéficiez d'une prise en charge avancée d'OpenStack et des technologies de conteneurs selon vos besoins. NetApp propose également des outils de sauvegarde basés sur le cloud (SnapMirror Cloud, Cloud Backup Service et Cloud Sync) et de hiérarchisation et d'archivage du stockage (FabricPool) pour ONTAP afin de réduire les dépenses d'exploitation et de tirer parti de la large portée du cloud.
- **Et plus.** tirez parti des performances extrêmes des baies NetApp AFF A-Series pour accélérer votre infrastructure virtualisée tout en gérant les coûts. Assurez la continuité totale de l'activité, qu'il s'agisse de la maintenance ou des mises à niveau, ou du remplacement complet de votre système de stockage à l'aide de clusters ONTAP scale-out. Protégez vos données au repos avec les fonctionnalités de chiffrement NetApp, sans frais supplémentaires. Assurez-vous que les performances respectent les niveaux de service grâce à des fonctionnalités de qualité de service très avancées. Elles font toutes partie du vaste éventail de fonctionnalités fournies par ONTAP, le logiciel de gestion des données d'entreprise leader du secteur.

Stockage unifié

ONTAP unifie le stockage selon une approche Software-defined simplifiée pour une gestion sécurisée et efficace, des performances améliorées et une évolutivité transparente. Cette approche améliore la protection des données et permet une utilisation efficace des ressources cloud.

À l'origine, cette approche unifiée faisait référence à la prise en charge des protocoles NAS et SAN sur un système de stockage unique. ONTAP continue d'être l'une des principales plateformes pour SAN, tout comme sa puissance initiale en matière de stockage NAS. ONTAP prend désormais également en charge le protocole objet S3. Bien que S3 ne soit pas utilisé pour les datastores, vous pouvez l'utiliser pour les applications hôtes. Pour en savoir plus sur la prise en charge du protocole S3 dans ONTAP"Présentation de la configuration S3", consultez le . Le terme stockage unifié a évolué pour signifier une approche unifiée de la gestion du stockage, notamment la possibilité de gérer toutes les ressources de stockage à partir d'une interface unique. Vous pouvez ainsi gérer à la fois les ressources de stockage sur site et dans le cloud, les derniers systèmes ASA, ainsi que plusieurs systèmes de stockage à partir d'une interface unique.

Une machine virtuelle de stockage (SVM) est l'unité de colocation sécurisée dans ONTAP. Il s'agit d'une structure logique permettant aux clients d'accéder aux systèmes exécutant ONTAP. Les SVM peuvent transmettre simultanément les données par le biais de plusieurs protocoles d'accès aux données via des interfaces logiques (LIF). Les SVM fournissent un accès aux données de niveau fichier via les protocoles NAS, tels que CIFS et NFS, et un accès aux données de niveau bloc via les protocoles SAN, tels que iSCSI, FC/FCoE et NVMe. Les SVM peuvent fournir des données aux clients SAN et NAS de façon indépendante et en même temps avec S3.



Dans le monde de vSphere, cette approche peut également se traduire par un système unifié d'infrastructure de postes de travail virtuels (VDI) avec une infrastructure de serveurs virtuels (VSI). Les systèmes qui exécutent ONTAP sont généralement moins onéreux pour VSI que les baies d'entreprise traditionnelles, tout en offrant des fonctionnalités avancées d'efficacité du stockage pour la gestion de l'infrastructure VDI dans le même système. ONTAP unifie également une grande variété de supports de stockage, des SSD aux SATA, et peut s'étendre facilement au cloud. Il n'est pas nécessaire d'acheter une baie Flash pour les performances, une baie SATA pour l'archivage ou des systèmes distincts pour le cloud. ONTAP les lie tous ensemble.

REMARQUE : pour plus d'informations sur les SVM, le stockage unifié et l'accès client, voir "Virtualisation du stockage" Dans le centre de documentation ONTAP 9.

Outils de virtualisation pour ONTAP

NetApp propose plusieurs outils logiciels autonomes compatibles avec les systèmes ONTAP et ASA classiques, permettant ainsi d'intégrer vSphere pour gérer efficacement votre environnement virtualisé.

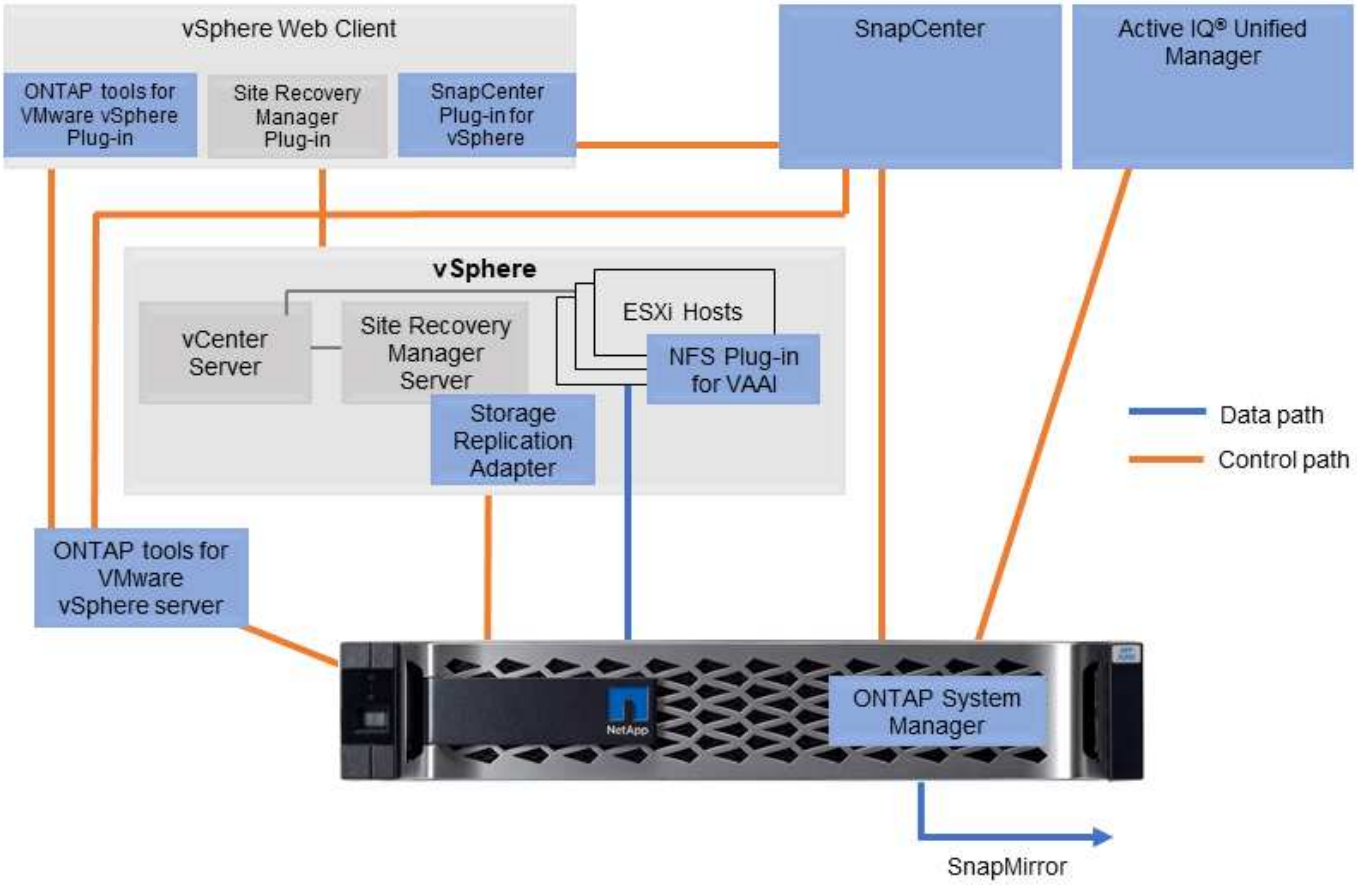
Les outils suivants sont inclus avec la licence ONTAP One sans frais supplémentaires. Voir la Figure 1 pour une description du fonctionnement de ces outils dans votre environnement vSphere.

Les outils ONTAP pour VMware vSphere

"Les outils ONTAP pour VMware vSphere" Est un ensemble d'outils permettant d'utiliser le stockage ONTAP avec vSphere. Le plug-in vCenter, précédemment appelé Virtual Storage Console (VSC), simplifie les fonctionnalités de gestion et d'efficacité du stockage, améliore la disponibilité et réduit les coûts de stockage ainsi que les charges opérationnelles, que vous utilisiez SAN ou NAS. Il s'appuie sur les bonnes pratiques pour le provisionnement des datastores et optimise les paramètres d'hôte ESXi pour les environnements de stockage NFS et bloc. Pour tous ces avantages, NetApp recommande d'utiliser ces outils ONTAP comme bonne pratique lors de l'utilisation de vSphere avec les systèmes exécutant ONTAP. Elle inclut une appliance de serveur, des extensions d'interface utilisateur pour vCenter, VASA Provider et Storage Replication adapter. La quasi-totalité des outils ONTAP peuvent être automatisés à l'aide d'API REST simples et consommables par la plupart des outils d'automatisation modernes.

- **Extensions de l'interface utilisateur vCenter.** Les extensions de l'interface utilisateur des outils ONTAP simplifient le travail des équipes opérationnelles et des administrateurs vCenter en intégrant des menus contextuels faciles à utiliser pour la gestion des hôtes et du stockage, des portlets informatifs et des fonctionnalités d'alerte natives directement dans l'interface utilisateur vCenter pour optimiser les workflows.
- **VASA Provider pour ONTAP.** le fournisseur VASA pour ONTAP prend en charge l'infrastructure VMware vStorage APIs for Storage Awareness (VASA). Il est fourni en tant qu'appliance virtuelle unique, avec les outils ONTAP pour VMware vSphere pour une facilité de déploiement. Vasa Provider connecte vCenter Server avec ONTAP pour faciliter le provisionnement et la surveillance du stockage des machines virtuelles. Il assure la prise en charge de VMware Virtual volumes (vvols), la gestion des profils de capacité de stockage et les performances individuelles de VM vvols, ainsi que des alarmes pour le contrôle de la capacité et de la conformité avec les profils.
- **Storage Replication adapter.** Le SRA est utilisé conjointement avec VMware Live site Recovery (VLSR)/site Recovery Manager (SRM) pour gérer la réplication des données entre les sites de production et de reprise après incident à l'aide de SnapMirror pour la réplication basée sur les baies. Grâce à elle, vous pouvez automatiser la tâche de basculement en cas d'incident et tester les répliques sans interruption pour une solution de reprise après incident fiable.

La figure suivante représente les outils ONTAP pour vSphere.



Plug-in SnapCenter pour VMware vSphere

Le "Plug-in SnapCenter pour VMware vSphere" est un plug-in pour vCenter Server qui vous permet de gérer les sauvegardes et les restaurations de machines virtuelles (VM) et de datastores. Cette solution fournit une interface unique pour la gestion des sauvegardes, des restaurations et des clones des machines virtuelles et des datastores sur plusieurs systèmes ONTAP. SnapCenter prend en charge la réplication vers et la restauration à partir de sites secondaires à l'aide de SnapMirror. Les dernières versions prennent également en charge SnapMirror dans le cloud (S3), les copies Snapshot inviolable, SnapLock et la synchronisation active SnapMirror. Le plug-in SnapCenter pour VMware vSphere peut être intégré aux plug-ins d'application SnapCenter afin d'assurer des sauvegardes cohérentes avec les applications.

Plug-in NFS pour VMware VAAI

Le "Plug-in NetApp NFS pour VMware VAAI" est un plug-in pour les hôtes ESXi qui leur permet d'utiliser les fonctionnalités VAAI avec les datastores NFS sur ONTAP. Il prend en charge le déchargement des copies pour les opérations de clonage, la réservation d'espace pour les fichiers de disque virtuel épais et le déchargement des snapshots. Le transfert des opérations de copie vers le stockage n'est pas forcément plus rapide. Toutefois, il réduit les besoins en bande passante réseau et réduit la charge des ressources hôte telles que les cycles de CPU, les tampons et les files d'attente. Vous pouvez utiliser les outils ONTAP pour VMware vSphere pour installer le plug-in sur des hôtes ESXi ou, le cas échéant, vSphere Lifecycle Manager (vLCM).

Options logicielles Premium

Les logiciels premium suivants sont disponibles auprès de NetApp. Ils ne sont pas inclus avec la licence ONTAP One et doivent être achetés séparément. * "Reprise après incident BlueXP " Pour VMware vSphere. Il s'agit d'un service cloud qui assure la reprise après incident et la sauvegarde des environnements VMware. Il peut être utilisé avec ou sans SnapCenter et prend en charge la reprise d'activité sur site à l'aide de SAN ou

NAS, ainsi que l'environnement sur site vers/depuis le cloud à l'aide de NFS, le cas échéant. * "Informations sur l'infrastructure de données (DII)". Il s'agit d'un service cloud pour la surveillance et l'analytique des environnements VMware. Elle prend en charge d'autres fournisseurs de stockage dans des environnements de stockage hétérogènes, ainsi que plusieurs fournisseurs de commutateurs et d'autres hyperviseurs. DII fournit des informations complètes sur les performances, la capacité et l'état de votre environnement VMware.

Volumes virtuels (vvols) et gestion basée sur des règles de stockage (SPBM)

Annoncé pour la première fois en 2012, NetApp a été l'un des premiers partenaires de conception avec VMware dans le développement de VMware vSphere APIs for Storage Awareness (VASA), la base de la gestion basée sur des règles de stockage (SPBM) avec des baies de stockage d'entreprise. Avec cette approche, la gestion du stockage granulaire des ordinateurs virtuels était limitée au stockage VMFS et NFS.

En tant que partenaire de conception technologique, NetApp a apporté son avis sur l'architecture et a annoncé en 2015 la prise en charge de vVols. Cette nouvelle technologie permet désormais d'automatiser le provisionnement du stockage granulaire au niveau des serveurs virtuels et véritablement natif des baies via la gestion du stockage basée sur des règles (SBPM).

Volumes virtuels (vVols)

Les vVols sont une architecture de stockage révolutionnaire qui permet la gestion granulaire du stockage des machines virtuelles. Le stockage peut ainsi être géré non seulement par machine virtuelle (y compris les métadonnées des machines virtuelles), mais également par VMDK. Les vVols sont un composant clé de la stratégie Software Defined Data Center (SDDC) qui constitue la base de VMware Cloud Foundation (VCF), fournissant ainsi une architecture de stockage plus efficace et évolutive pour les environnements virtualisés.

Les vVols permettent aux machines virtuelles de consommer du stockage par machine virtuelle, car chaque objet de stockage de machine virtuelle est une entité unique dans NetApp ONTAP. Avec les systèmes ASA r2 qui ne nécessitent plus de gestion de volume, chaque objet de stockage VM est une unité de stockage unique sur la baie et peut être contrôlé de manière indépendante. Cela permet de créer des règles de stockage qui peuvent être appliquées aux machines virtuelles individuelles ou aux VMDK (et ainsi aux unités d'exploitation doubles), fournissant un contrôle granulaire sur les services de stockage tels que les performances, la disponibilité et la protection des données.

Gestion du stockage basée sur des règles (SBPM)

Grâce à la gestion du stockage basée sur des règles, une structure sert de couche d'abstraction entre les services de stockage disponibles pour votre environnement de virtualisation et les éléments de stockage provisionnés via des règles. Cette approche permet aux architectes du stockage de concevoir des pools de stockage avec des fonctionnalités différentes. Ces pools peuvent être facilement consommés par les administrateurs des VM. Les administrateurs peuvent ensuite faire correspondre les besoins des charges de travail des machines virtuelles aux pools de stockage provisionnés. Cette approche simplifie la gestion du stockage et permet une utilisation plus efficace des ressources de stockage.

Le SBPM est un composant clé des vVols qui fournit un framework basé sur des règles pour la gestion des services de stockage. Les règles sont créées par les administrateurs vSphere à l'aide de règles et de fonctionnalités exposées par le VASA Provider(VP) du fournisseur. Il est possible de créer des règles pour différents services de stockage, tels que les performances, la disponibilité et la protection des données. Il est possible d'attribuer des règles à des machines virtuelles ou des VMDK individuels pour assurer un contrôle granulaire des services de stockage.

NetApp ONTAP et vVols

NetApp ONTAP est leader du secteur du stockage en vVols à l'échelle du cluster, prenant en charge des centaines de milliers de vVols* par cluster unique. En revanche, les fournisseurs de baies d'entreprise et de baies Flash plus petites prennent en charge jusqu'à plusieurs milliers de vVols par baie. ONTAP offre une solution de stockage évolutive et efficace pour les environnements VMware vSphere, prenant en charge les vVols avec un ensemble complet de services de stockage, dont la déduplication, la compression, le provisionnement fin et la protection des données. La gestion du stockage basée sur des règles facilite l'intégration transparente aux environnements VMware vSphere.

Nous avons mentionné précédemment que les administrateurs des ordinateurs virtuels peuvent consommer de la capacité sous forme de pools de stockage. Pour ce faire, nous utilisons des conteneurs de stockage représentés dans vSphere en tant que datastores logiques.

Les conteneurs de stockage sont créés par les administrateurs du stockage et servent à grouper les ressources de stockage consommées par les administrateurs des VM. Les conteneurs de stockage peuvent être créés différemment en fonction du type de système ONTAP que vous utilisez. Avec les clusters ONTAP 9 classiques, les conteneurs se voient attribuer un ou plusieurs volumes FlexVol qui forment le pool de stockage. Avec les systèmes ASA r2, l'intégralité du cluster correspond au pool de stockage.

> Pour plus d'informations sur les volumes virtuels VMware vSphere, SPBM et ONTAP, voir "Tr-4400 : volumes virtuels VMware vSphere avec ONTAP".

*Selon la plate-forme et le protocole

Datastores et protocoles

Présentation des fonctionnalités de datastore et de protocole vSphere

Six protocoles sont utilisés pour connecter VMware vSphere aux datastores d'un système exécutant ONTAP :

- FCP
- NVMe/FC
- NVMe/TCP
- ISCSI
- NFS v3
- NFS v4.1

FCP, NVMe/FC, NVMe/TCP et iSCSI sont des protocoles en mode bloc qui utilisent VMFS (Virtual machine File System) vSphere pour stocker des machines virtuelles dans des LUN ONTAP ou des namespaces NVMe contenus dans un ONTAP FlexVol volume. NFS est un protocole de fichier qui place les machines virtuelles dans des datastores (qui sont simplement des volumes ONTAP) sans avoir besoin de VMFS. SMB (CIFS), iSCSI, NVMe/TCP ou NFS peuvent également être utilisés directement d'un système d'exploitation invité à ONTAP.

Les tableaux suivants présentent les fonctionnalités de datastore traditionnel prises en charge par vSphere avec ONTAP. Ces informations ne s'appliquent pas aux datastores vvols, mais elles s'appliquent généralement aux versions vSphere 6.x et ultérieures utilisant des versions ONTAP prises en charge. Vous pouvez également consulter le "Outil VMware Configuration Maximums" pour connaître les versions de vSphere spécifiques afin de confirmer les limites spécifiques.

Capacités/fonctionnalités	FC	ISCSI	NVMe-of	NFS
Format	Mappage de périphériques VMFS ou bruts (RDM)	VMFS ou RDM	VMFS	s/o
Nombre maximal de datastores ou de LUN	1024 LUN par hôte	1024 LUN par serveur	256 Namespeces par serveur	256 connexions NFS par hôte (impactées par nconnect et session Trunking) NFS par défaut. MaxVolumes est 8. Utilisez les outils ONTAP pour VMware vSphere et augmentez jusqu'à 256.
Taille maximale des datastores	64 TO	64 TO	64 TO	FlexVol volume 300 To ou plus avec un volume FlexGroup
Taille maximale des fichiers du datastore	62TO	62TO	62TO	62 To avec ONTAP 9.12.1P2 et versions ultérieures
Profondeur de file d'attente optimale par LUN ou par système de fichiers	64-256	64-256	Négociation automatique	Se reporter à NFS.MaxQueueDepth dans "Hôte ESXi recommandé et autres paramètres ONTAP recommandés".

Le tableau suivant répertorie les fonctionnalités de stockage VMware prises en charge.

Capacité/fonctionnalité	FC	ISCSI	NVMe-of	NFS
VMotion	Oui.	Oui.	Oui.	Oui.
Stockage vMotion	Oui.	Oui.	Oui.	Oui.
Haute disponibilité VMware	Oui.	Oui.	Oui.	Oui.
Storage Distributed Resource Scheduler (SDRS)	Oui.	Oui.	Oui.	Oui.
Logiciel de sauvegarde compatible VADP (VMware vStorage APIs for Data protection)	Oui.	Oui.	Oui.	Oui.

Capacité/fonctionnalité	FC	ISCSI	NVMe-of	NFS
Microsoft Cluster Service (MSCS) ou mise en cluster de basculement au sein d'une machine virtuelle	Oui.	Oui [1]	Oui [1]	Non pris en charge
Tolérance aux pannes	Oui.	Oui.	Oui.	Oui.
Live site Recovery Manager/site Recovery Manager	Oui.	Oui.	Non [2]	V3 uniquement [2]
Machines virtuelles à provisionnement fin (disques virtuels)	Oui.	Oui.	Oui.	Oui. Ce paramètre est le paramètre par défaut pour toutes les machines virtuelles sur NFS lorsqu'elles n'utilisent pas VAAI.
Chemins d'accès multiples natifs VMware	Oui.	Oui.	Oui.	L'agrégation de sessions NFS v4.1 requiert ONTAP 9.14.1 et versions ultérieures

Le tableau suivant répertorie les fonctionnalités de gestion du stockage ONTAP prises en charge.

Capacités/fonctionnalités	FC	ISCSI	NVMe-of	NFS
Déduplication des données	D'économies sur la baie	D'économies sur la baie	D'économies sur la baie	Économies au niveau du datastore
Provisionnement fin	Datastore ou RDM	Datastore ou RDM	Datastore	Datastore
Redimensionnement datastore	Évoluer uniquement	Évoluer uniquement	Évoluer uniquement	Croissance, croissance automatique et réduction des volumes
Plug-ins SnapCenter pour applications Windows, Linux (invités)	Oui.	Oui.	Oui.	Oui.
Contrôle et configuration de l'hôte à l'aide des outils ONTAP pour VMware vSphere	Oui.	Oui.	Oui.	Oui.

Capacités/fonctionnalités	FC	ISCSI	NVMe-of	NFS
Provisionnement avec les outils ONTAP pour VMware vSphere	Oui.	Oui.	Oui.	Oui.

Le tableau suivant répertorie les fonctionnalités de sauvegarde prises en charge.

Capacités/fonctionnalités	FC	ISCSI	NVMe-of	NFS
Snapshots ONTAP	Oui.	Oui.	Oui.	Oui.
SRM pris en charge par les sauvegardes répliquées	Oui.	Oui.	Non [2]	V3 uniquement [2]
SnapMirror volume	Oui.	Oui.	Oui.	Oui.
Accès image VMDK	Logiciels de sauvegarde compatibles SnapCenter et VADP	Logiciels de sauvegarde compatibles SnapCenter et VADP	Logiciels de sauvegarde compatibles SnapCenter et VADP	Logiciel de sauvegarde compatible SnapCenter et VADP, client vSphere et navigateur de datastore du client Web vSphere
Accès niveau fichier VMDK	Logiciel de sauvegarde compatible SnapCenter et VADP, Windows uniquement	Logiciel de sauvegarde compatible SnapCenter et VADP, Windows uniquement	Logiciel de sauvegarde compatible SnapCenter et VADP, Windows uniquement	Logiciels de sauvegarde compatibles SnapCenter et VADP et applications tierces
Granularité NDMP	Datastore	Datastore	Datastore	Datastore ou VM

[1] **NetApp recommande** d'utiliser iSCSI dans l'invité pour les clusters Microsoft plutôt que des VMDK compatibles avec les enregistreurs multiples dans un datastore VMFS. Cette approche est entièrement prise en charge par Microsoft et VMware. Elle offre une grande flexibilité avec ONTAP (SnapMirror vers les systèmes ONTAP sur site ou dans le cloud), est facile à configurer et à automatiser, et peut être protégée avec SnapCenter. VSphere 7 ajoute une nouvelle option VMDK en cluster. Ceci est différent des VMDK compatibles avec le multiwriter, qui nécessitent un datastore VMFS 6 dont la prise en charge des VMDK en cluster est activée. D'autres restrictions s'appliquent. Consultez la documentation de VMware "Configuration de Windows Server Failover Clustering"pour obtenir des instructions de configuration.

Les datastores [2] utilisant NVMe-of et NFS v4.1 requièrent une réplication vSphere. La réplication basée sur les baies pour NFS v4.1 n'est pas actuellement prise en charge par SRM. La réplication basée sur la baie avec NVMe-of n'est actuellement pas prise en charge par l'outil ONTAP pour VMware vSphere Storage Replication adapter (SRA).

Sélection d'un protocole de stockage

Les systèmes exécutant ONTAP prennent en charge les principaux protocoles de stockage. Les clients peuvent ainsi choisir l'environnement le mieux adapté à leur environnement, en fonction de l'infrastructure réseau existante et planifiée, et des compétences du personnel. Historiquement, les tests NetApp ont généralement montré peu de différence entre les protocoles s'exécutant à des vitesses de ligne similaires et le nombre de connexions. Cependant, NVMe-of (NVMe/TCP et NVMe/FC) montre des gains remarquables en matière d'IOPS, de réduction de la latence et de réduction d'au moins 50 % de la consommation du processeur hôte par les E/S de stockage. À l'autre extrémité du spectre, NFS offre une flexibilité et une facilité de gestion optimales, en particulier pour un grand nombre de machines virtuelles. Tous ces protocoles peuvent être utilisés et gérés avec les outils ONTAP pour VMware vSphere, qui offrent une interface simple pour créer et gérer des datastores.

Les facteurs suivants peuvent être utiles lors de l'examen d'un choix de protocole :

- **Environnement de fonctionnement actuel.** Bien que les équipes INFORMATIQUES soient généralement compétentes en matière de gestion de l'infrastructure IP Ethernet, elles ne sont pas toutes compétentes en matière de gestion d'une structure SAN FC. Cependant, l'utilisation d'un réseau IP générique non conçu pour le trafic de stockage risque de ne pas fonctionner correctement. Considérez l'infrastructure de réseau que vous avez en place, toutes les améliorations planifiées, ainsi que les compétences et la disponibilité du personnel pour les gérer.
- **Simplicité d'installation.** au-delà de la configuration initiale de la structure FC (commutateurs et câblage supplémentaires, segmentation et vérification de l'interopérabilité des HBA et des micrologiciels), les protocoles de bloc exigent également la création et le mappage de LUN, ainsi que la découverte et le formatage par le système d'exploitation invité. Une fois les volumes NFS créés et exportés, ils sont montés par l'hôte ESXi et prêts à être utilisés. Avec NFS, il n'a pas de qualification de matériel ni de firmware à gérer.
- * Facilité de gestion.* Avec les protocoles SAN, si vous avez besoin de plus d'espace, plusieurs étapes sont nécessaires, notamment la croissance d'un LUN, la nouvelle analyse pour découvrir la nouvelle taille, puis l'expansion du système de fichiers). Bien que la croissance d'une LUN soit possible, la réduction de sa taille n'est pas le cas. NFS facilite le dimensionnement et le redimensionnement peut être automatisé par le système de stockage. SAN offre une récupération d'espace via les commandes DEALLOCATE/TRIM/UNMAP du système d'exploitation invité, ce qui permet de renvoyer dans la baie l'espace des fichiers supprimés. Ce type de récupération d'espace n'est pas difficile avec les datastores NFS.
- **Transparence de l'espace de stockage.** l'utilisation du stockage est généralement plus facile à voir dans les environnements NFS parce que le provisionnement fin renvoie immédiatement des économies. De même, les économies de déduplication et de clonage sont immédiatement disponibles pour les autres VM dans le même datastore ou pour les autres volumes du système de stockage. La densité des machines virtuelles est également meilleure généralement dans un datastore NFS, ce qui permet d'améliorer les économies de déduplication et de réduire les coûts de gestion en utilisant moins de datastores à gérer.

Disposition des datastores

Les systèmes de stockage ONTAP offrent une grande flexibilité de création de datastores pour les machines virtuelles et les disques virtuels. Bien que de nombreuses bonnes pratiques ONTAP soient appliquées lors de l'utilisation des outils ONTAP pour provisionner des datastores pour vSphere (répertoriés dans la section "Hôte ESXi recommandé et autres paramètres ONTAP recommandés"), voici quelques instructions supplémentaires à prendre en compte :

- Le déploiement de vSphere avec des datastores NFS ONTAP offre une implémentation très performante et facile à gérer qui fournit des ratios VM/datastore qui ne peuvent pas être obtenus avec des protocoles de stockage de niveau bloc. Cette architecture peut entraîner une multiplication par dix de la densité des datastores avec une corrélation réduction du nombre de datastores. Bien qu'un datastore plus volumineux

puisse bénéficier de l'efficacité du stockage et offrir des avantages opérationnels, envisagez d'utiliser au moins quatre datastores (volumes FlexVol) par nœud pour stocker vos machines virtuelles sur un seul contrôleur ONTAP afin d'optimiser les performances des ressources matérielles. Cette approche vous permet également de créer des datastores avec différentes règles de restauration. Certaines peuvent être sauvegardées ou répliquées plus fréquemment que d'autres, en fonction des besoins de l'entreprise. Les volumes FlexGroup n'ont pas besoin de plusieurs datastores pour améliorer les performances, car ils évoluent indépendamment de la conception.

- **NetApp recommande** l'utilisation de volumes FlexVol pour la plupart des datastores NFS. À partir de ONTAP 9.8, les volumes FlexGroup sont également pris en charge en tant que datastores et sont généralement recommandés pour certaines utilisations. Les autres conteneurs de stockage ONTAP, tels que les qtrees, ne sont généralement pas recommandés, car ils ne sont actuellement pas pris en charge par les outils ONTAP pour VMware vSphere ou par le plug-in NetApp SnapCenter pour VMware vSphere.
- La taille correcte des datastores de volumes FlexVol est d'environ 4 To à 8 To. Cette taille constitue un bon équilibre pour les performances, la facilité de gestion et la protection des données. Démarrer petit (disons 4 To) et développer le datastore en fonction des besoins (jusqu'au maximum 300 To) Les datastores plus petits peuvent être plus rapides à restaurer depuis la sauvegarde ou après un incident, et déplacés rapidement dans l'ensemble du cluster. Envisagez d'utiliser la fonction de dimensionnement automatique de ONTAP pour augmenter et réduire automatiquement le volume en fonction des modifications de l'espace utilisé. L'assistant ONTAP Tools for VMware vSphere datastore Provisioning utilise le dimensionnement automatique par défaut pour les nouveaux datastores. Vous pouvez également personnaliser davantage les seuils d'extension et de réduction ainsi que la taille maximale et minimale, avec System Manager ou la ligne de commandes.
- Les datastores VMFS peuvent également être configurés avec des LUN ou des espaces de noms NVMe (appelés unités de stockage dans les nouveaux systèmes ASA) accessibles via FC, iSCSI, NVMe/FC ou NVMe/TCP. VMFS permet à chaque serveur ESX d'un cluster d'accéder simultanément aux datastores. Les datastores VMFS peuvent être jusqu'à 64 To et comprennent jusqu'à 32 LUN de 2 To (VMFS 3) ou un seul LUN de 64 To (VMFS 5). La taille de LUN maximale de la baie ONTAP est de 128 To sur les systèmes AFF, ASA et FAS. NetApp recommande toujours d'utiliser une LUN unique et volumineuse pour chaque datastore, plutôt que d'utiliser les extensions. Comme pour NFS, envisagez d'utiliser plusieurs datastores (volumes ou unités de stockage) pour optimiser les performances sur un seul contrôleur ONTAP.
- Les anciens systèmes d'exploitation invités (OS) devaient s'aligner sur le système de stockage pour obtenir des performances et une efficacité du stockage optimales. Cependant, les systèmes d'exploitation actuels pris en charge par les fournisseurs de Microsoft et de distributeurs Linux tels que Red Hat ne nécessitent plus d'ajustements pour aligner la partition du système de fichiers sur les blocs du système de stockage sous-jacent dans un environnement virtuel. Si vous utilisez un ancien système d'exploitation qui peut nécessiter un alignement, recherchez dans la base de connaissances du support NetApp des articles « alignement des machines virtuelles » ou demandez une copie de l'article TR-3747 à un contact partenaire ou commercial NetApp.
- Évitez d'utiliser des utilitaires de défragmentation au sein du système d'exploitation invité, car cela n'améliore pas les performances et affecte l'efficacité du stockage et l'utilisation de l'espace Snapshot. Envisagez également de désactiver l'indexation des recherches sur le système d'exploitation invité pour les postes de travail virtuels.
- ONTAP s'est leader du marché en proposant des fonctionnalités innovantes d'efficacité du stockage qui vous permettent d'exploiter au maximum votre espace disque utilisable. Les systèmes AFF renforcent cette efficacité avec la compression et la déduplication à la volée par défaut. Les données sont dédupliquées sur tous les volumes d'un agrégat. Ainsi, vous n'avez plus besoin de regrouper des systèmes d'exploitation similaires et des applications similaires au sein d'un même datastore pour optimiser les économies.
- Dans certains cas, vous n'aurez même pas besoin d'un datastore. Envisagez des systèmes de fichiers invités, tels que NFS, SMB, NVMe/TCP ou iSCSI gérés par l'invité. Pour une assistance spécifique aux applications, consultez les rapports techniques de NetApp pour votre application. Par exemple"Les bases

de données Oracle sur ONTAP", a une section sur la virtualisation avec des détails utiles.

- Les disques de première classe (ou des disques virtuels améliorés) permettent de gérer des disques gérés par vCenter indépendamment d'une machine virtuelle dotée de vSphere 6.5 et versions ultérieures. Lorsqu'elles sont principalement gérées par API, elles peuvent être utiles avec vvols, en particulier lorsqu'elles sont gérées par les outils OpenStack ou Kubernetes. Ils sont pris en charge par ONTAP ainsi que par les outils ONTAP pour VMware vSphere.

Migration des datastores et des machines virtuelles

Lorsque vous migrez des machines virtuelles depuis un datastore existant sur un autre système de stockage vers ONTAP, voici quelques principes à prendre en compte :

- Utilisez Storage vMotion pour déplacer la masse de vos machines virtuelles vers ONTAP. Cette approche n'assure pas seulement une exécution sans interruption des machines virtuelles. Elle permet également d'exploiter des fonctionnalités d'efficacité du stockage de ONTAP, comme la déduplication et la compression à la volée, pour traiter les données lors de leur migration. Envisagez d'utiliser les fonctionnalités de vCenter pour sélectionner plusieurs machines virtuelles dans la liste d'inventaire, puis planifiez la migration (utilisez la touche Ctrl tout en cliquant sur actions) à un moment opportun.
- Bien que vous puissiez planifier avec soin une migration vers des datastores de destination appropriés, il est souvent plus simple de les migrer en bloc, puis de les organiser ultérieurement, si nécessaire. Utilisez cette approche pour orienter la migration vers différents datastores si vous avez besoin de protection des données spécifique, par exemple des calendriers Snapshot différents. De plus, une fois les VM sur le cluster NetApp, le stockage vMotion peut utiliser les délestages VAAI pour déplacer des VM entre les datastores du cluster sans nécessiter de copie basée sur l'hôte. Notez que NFS ne décharge pas le stockage vMotion des machines virtuelles optimisées, mais VMFS.
- Les machines virtuelles qui nécessitent une migration plus minutieuse incluent les bases de données et les applications qui utilisent le stockage associé. De manière générale, envisagez l'utilisation des outils de l'application pour gérer la migration. Pour Oracle, envisagez d'utiliser des outils Oracle tels que RMAN ou ASM pour migrer les fichiers de base de données. Voir "Migration des bases de données Oracle vers des systèmes de stockage ONTAP" pour plus d'informations. De même, pour SQL Server, envisagez d'utiliser soit SQL Server Management Studio, soit des outils NetApp tels qu'SnapManager pour SQL Server, soit SnapCenter.

Les outils ONTAP pour VMware vSphere

La meilleure pratique la plus importante lors de l'utilisation de vSphere avec des systèmes exécutant ONTAP consiste à installer et à utiliser le plug-in ONTAP Tools for VMware vSphere (anciennement Virtual Storage Console). Ce plug-in vCenter simplifie la gestion du stockage, améliore la disponibilité et réduit les coûts de stockage et la surcharge opérationnelle, que ce soit avec SAN ou NAS, sur ASA, AFF, FAS ou même ONTAP Select (une version Software-defined ONTAP exécutée sur une machine virtuelle VMware ou KVM). Il tire parti des bonnes pratiques pour le provisionnement des datastores et optimise les paramètres des hôtes ESXi pour les délais entre les chemins d'accès multiples et les HBA (ces paramètres sont décrits dans l'annexe B). Comme il s'agit d'un plug-in vCenter, il est disponible pour tous les clients Web vSphere qui se connectent au serveur vCenter.

Le plug-in permet également d'utiliser d'autres outils ONTAP dans les environnements vSphere. Il vous permet d'installer le plug-in NFS pour VMware VAAI, ce qui permet d'alléger la copie vers ONTAP pour les opérations de clonage de machines virtuelles, de réserver de l'espace pour les fichiers de disques virtuels lourds et de décharger les snapshots ONTAP.

> Sur les clusters vSphere basés sur image, vous voulez toujours ajouter le plug-in NFS à votre image afin qu'ils ne soient pas hors conformité lors de l'installation avec les outils ONTAP.

Les outils ONTAP sont également l'interface de gestion de nombreuses fonctions du fournisseur VASA pour ONTAP, prenant en charge la gestion basée sur des règles de stockage avec vVols.

En général, **NetApp recommande** d'utiliser les outils ONTAP pour l'interface VMware vSphere dans vCenter pour provisionner les datastores traditionnels et vVols afin de s'assurer du respect des bonnes pratiques.

Réseau général

La configuration des paramètres réseau lors de l'utilisation de vSphere avec des systèmes exécutant ONTAP est simple et similaire à celle des autres configurations réseau. Voici quelques points à prendre en compte :

- Trafic du réseau de stockage séparé des autres réseaux Un réseau distinct peut être obtenu à l'aide d'un VLAN dédié ou de commutateurs distincts pour le stockage. Si le réseau de stockage partage des chemins physiques, tels que des liaisons ascendantes, vous pouvez avoir besoin de la qualité de service ou de ports supplémentaires pour garantir une bande passante suffisante. Ne connectez pas les hôtes directement au stockage ; utilisez les commutateurs pour disposer de chemins redondants et permettez à VMware HA de fonctionner sans intervention. Voir "Connexion directe au réseau" pour plus d'informations.
- Les trames Jumbo peuvent être utilisées si vous le souhaitez et prises en charge par votre réseau, en particulier lors de l'utilisation d'iSCSI. Si elles sont utilisées, assurez-vous qu'elles sont configurées de manière identique sur tous les périphériques réseau, VLAN, etc. Dans le chemin entre le stockage et l'hôte ESXi. Vous pourriez voir des problèmes de performances ou de connexion. La MTU doit également être définie de manière identique sur le switch virtuel ESXi, le port VMkernel et également sur les ports physiques ou les groupes d'interface de chaque nœud ONTAP.
- NetApp recommande uniquement de désactiver le contrôle de flux réseau sur les ports d'interconnexion de cluster au sein d'un cluster ONTAP. NetApp ne recommande pas d'autres recommandations sur les meilleures pratiques pour les ports réseau restants utilisés pour le trafic de données. Vous devez activer ou désactiver si nécessaire. Voir "TR-4182" pour plus d'informations sur le contrôle de flux.
- Lorsque les baies de stockage VMware ESXi et ONTAP sont connectées aux réseaux de stockage Ethernet, **NetApp recommande** de configurer les ports Ethernet auxquels ces systèmes se connectent en tant que ports de périphérie RSTP (Rapid Spanning Tree Protocol) ou en utilisant la fonction PortFast de Cisco. **NetApp recommande** d'activer la fonctionnalité Spanning-Tree PortFast trunk dans les environnements qui utilisent la fonctionnalité Cisco PortFast et dont l'agrégation VLAN 802.1Q est activée sur le serveur VMware ESXi ou sur les baies de stockage ONTAP.
- **NetApp recommande** les meilleures pratiques suivantes pour l'agrégation de liens :
  - Utilisez des commutateurs qui prennent en charge l'agrégation de liens des ports sur deux châssis de commutateurs distincts grâce à une approche de groupe d'agrégation de liens multichâssis, telle que Virtual PortChannel (VPC) de Cisco.
  - Désactiver LACP pour les ports de switch connectés à ESXi, sauf si vous utilisez dvswitches 5.1 ou version ultérieure avec LACP configuré.
  - Utilisez LACP pour créer des agrégats de liens pour les systèmes de stockage ONTAP avec des groupes d'interfaces multimode dynamiques avec un hachage de port ou d'IP. Reportez-vous à la section "Gestion de réseau" pour obtenir des conseils supplémentaires.
  - Utilisez une stratégie de regroupement de hachage IP sur ESXi lors de l'agrégation de liens statiques (EtherChannel, par exemple) et des vSwitch standard ou de l'agrégation de liens basée sur LACP avec des commutateurs distribués vSphere. Si l'agrégation de liens n'est pas utilisée, utilisez plutôt « route basée sur l'ID de port virtuel d'origine ».

SAN (FC, FCoE, NVMe/FC, iSCSI), RDM

Il existe quatre méthodes pour utiliser les périphériques de stockage en mode bloc dans vSphere :

- Avec les datastores VMFS
- Avec mappage de périphériques bruts (RDM)
- En tant que LUN connectée iSCSI ou espace de noms connecté à NVMe/TCP, accessible et contrôlé par un initiateur logiciel à partir d'un système d'exploitation invité de machine virtuelle
- Comme datastore vVols

VMFS est un système de fichiers en cluster hautes performances qui fournit des datastores sous forme de pools de stockage partagés. Les datastores VMFS peuvent être configurés avec des LUN accessibles via FC, iSCSI, FCoE ou avec des espaces de noms NVMe accessibles via les protocoles NVMe/FC ou NVMe/TCP. VMFS permet à chaque serveur ESX d'un cluster d'accéder simultanément au stockage. La taille de LUN maximale est généralement de 128 To à partir de ONTAP 9.12.1P2 (et versions antérieures avec les systèmes ASA). Par conséquent, un datastore VMFS 5 ou 6 de 64 To de taille maximale peut être créé à l'aide d'une seule LUN.

> Les extensions sont un concept de stockage vSphere dans lequel vous pouvez « assembler » plusieurs LUN pour créer un seul datastore plus grand. Vous ne devez jamais utiliser d'extensions pour atteindre la taille de datastore souhaitée. Une seule LUN est la meilleure pratique pour un datastore VMFS.

VSphere inclut la prise en charge intégrée de plusieurs chemins vers les périphériques de stockage. VSphere peut détecter le type de périphérique de stockage pour les systèmes de stockage pris en charge et configurer automatiquement la pile de chemins d'accès multiples afin de prendre en charge les fonctionnalités du système de stockage utilisé, la royldesse du protocole utilisé ou si ONTAP, AFF, FAS ou Software Defined ASA est utilisé.

VSphere et ONTAP prennent en charge ALUA (Asymmetric Logical Unit Access) pour établir des chemins actifs/optimisés et actifs/non optimisés pour Fibre Channel et iSCSI, et ANA (Asymmetric Namespace Access) pour les namespaces NVMe à l'aide de NVMe/FC et NVMe/TCP. Dans ONTAP, un chemin optimisé pour le protocole ALUA ou ANA suit un chemin d'accès direct aux données en utilisant un port cible sur le nœud qui héberge la LUN ou l'espace de noms auquel vous accédez. ALUA/ANA est activé par défaut dans vSphere et ONTAP. Le logiciel de chemins d'accès multiples de vSphere reconnaît le cluster ONTAP en tant qu'ALUA ou ANA et il utilise le plug-in natif approprié avec la règle d'équilibrage de charge round Robin.

Avec les systèmes ASA de NetApp, les LUN et les namespaces sont présentés aux hôtes ESXi avec des chemins d'accès symétriques. Ce qui signifie que tous les chemins sont actifs et optimisés. Le logiciel de chemins d'accès multiples de vSphere reconnaît le système ASA comme symétrique et utilise le plug-in natif approprié avec la règle d'équilibrage de charge round Robin.

> Reportez-vous à la section "Hôte ESXi recommandé et autres paramètres ONTAP recommandés" pour les paramètres de chemins d'accès multiples optimisés.

ESXi ne voit pas les LUN, les espaces de noms ou les chemins au-delà de ses limites. Dans un cluster ONTAP plus grand, il est possible d'atteindre la limite de chemin avant la limite de LUN. Pour résoudre cette limitation, ONTAP prend en charge le mappage de LUN sélectif (SLM) dans la version 8.3 et les versions ultérieures.

> Reportez-vous au "Outil VMware Configuration Maximums"pour connaître les limites les plus récentes prises en charge dans ESXi.

SLM limite les nœuds qui annoncent les chemins vers une LUN donnée. Il est recommandé pour NetApp d'avoir au moins deux LIF par nœud et par SVM et d'utiliser SLM pour limiter les chemins annoncés au nœud hébergeant la LUN et son partenaire HA. Bien que d'autres chemins existent, ils ne sont pas annoncés par

défaut. Il est possible de modifier les chemins annoncés avec les arguments de noeud de rapport ajouter et supprimer dans SLM. Notez que les LUN créées dans les versions antérieures à la version 8.3 annoncent tous les chemins et doivent être modifiés pour uniquement annoncer les chemins d'accès à la paire HA d'hébergement. Pour plus d'informations sur SLM, consultez la section 5.9 de "TR-4080". La méthode précédente de ensembles de ports peut également être utilisée pour réduire davantage les chemins disponibles pour une LUN. Les jeux de ports permettent de réduire le nombre de chemins visibles via lesquels les initiateurs d'un groupe initiateur peuvent voir les LUN.

- SLM est activé par défaut. Sauf si vous utilisez des ensembles de ports, aucune configuration supplémentaire n'est requise.
- Pour les LUN créées avant Data ONTAP 8.3, appliquez manuellement SLM en exécutant la `lun mapping remove-reporting-nodes` commande pour supprimer les nœuds de reporting de LUN et limiter l'accès de LUN au nœud propriétaire de la LUN et à son partenaire HA.

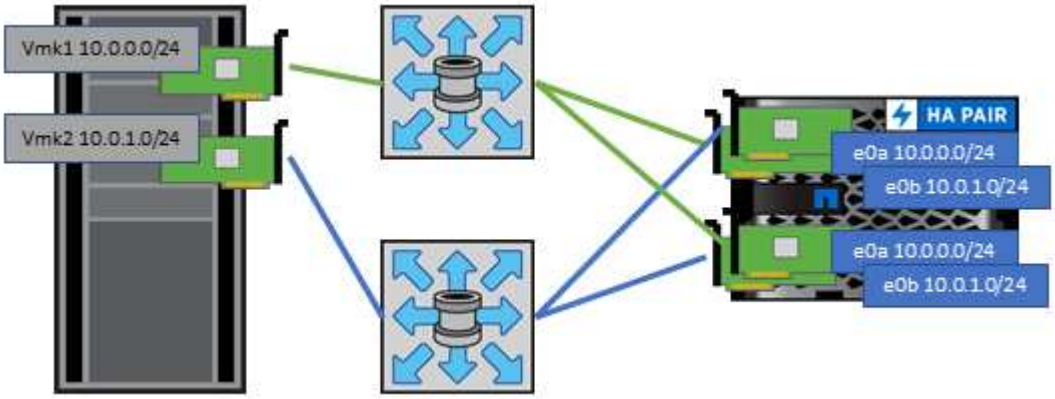
Les protocoles de blocs basés sur SCSI (iSCSI, FC et FCoE) accèdent aux LUN via des ID de LUN, des numéros de série et des noms uniques. Les protocoles FC et FCoE utilisent des noms WWN et WWPN (WWN) et iSCSI utilise des noms qualifiés iSCSI (IQN) pour établir des chemins basés sur les mappages de LUN à groupe initiateur filtrés par port et SLM. Pour gérer les protocoles de niveau bloc basés sur NVMe, il faut attribuer le namespace avec un ID d'espace de noms généré automatiquement à un sous-système NVMe, puis mapper ce sous-système sur le nom qualifié NVMe (NQN) du ou des hôtes. Indépendamment du FC ou du TCP, les namespaces NVMe sont mappés à l'aide du NQN et non du WWPN ou du WWNN. L'hôte crée ensuite un contrôleur défini par logiciel pour que le sous-système mappé puisse accéder à ses espaces de noms. Le chemin d'accès aux LUN et aux espaces de noms au sein de ONTAP n'a aucun sens pour les protocoles en mode bloc et n'est présenté nulle part dans le protocole. Par conséquent, un volume contenant uniquement des LUN n'a pas besoin d'être monté en interne et un chemin de jonction n'est pas nécessaire pour les volumes contenant les LUN utilisées dans les datastores.

D'autres meilleures pratiques à prendre en compte :

- Vérifiez "Hôte ESXi recommandé et autres paramètres ONTAP recommandés" les paramètres recommandés par NetApp en collaboration avec VMware.
- Vérifier qu'une interface logique (LIF) est créée pour chaque SVM sur chaque nœud du cluster ONTAP pour optimiser la disponibilité et la mobilité. La meilleure pratique du SAN de ONTAP est d'utiliser deux ports physiques et LIF par nœud, un pour chaque structure. ALUA sert à analyser les chemins et à identifier les chemins (directs) optimisés actifs/actifs au lieu de chemins non optimisés actifs. ALUA est utilisé pour FC, FCoE et iSCSI.
- Pour les réseaux iSCSI, utilisez plusieurs interfaces réseau VMkernel sur différents sous-réseaux du réseau avec le regroupement de cartes réseau lorsque plusieurs commutateurs virtuels sont présents. Vous pouvez également utiliser plusieurs cartes réseau physiques connectées à plusieurs commutateurs physiques pour fournir la haute disponibilité et un débit accru. La figure suivante fournit un exemple de connectivité multivoie. Dans ONTAP, configurez soit un groupe d'interface en mode unique pour basculement avec deux liaisons ou plus connectées à deux ou plusieurs switchs, soit au moyen de LACP ou d'une autre technologie d'agrégation de liens avec des groupes d'interfaces multimode afin d'assurer la haute disponibilité et les avantages de l'agrégation de liens.
- Si le protocole CHAP (Challenge-Handshake Authentication Protocol) est utilisé dans ESXi pour l'authentification de la cible, il doit également être configuré dans ONTAP à l'aide de l'interface de ligne de commande (`vserver iscsi security create`) Ou avec System Manager (modifier la sécurité de l'initiateur sous Storage > SVM > SVM Settings > protocoles > iSCSI).
- Utilisez les outils ONTAP pour VMware vSphere pour créer et gérer des LUN et des igroups. Le plug-in détermine automatiquement les WWPN des serveurs et crée les igroups appropriés. Il configure également les LUN en fonction des meilleures pratiques et les mappe avec les groupes initiateurs appropriés.
- Utilisez les RDM avec soin car ils peuvent être plus difficiles à gérer et ils utilisent également des chemins,

qui sont limités comme décrit précédemment. Les LUN ONTAP prennent en charge les deux "mode de compatibilité physique et virtuelle" RDM.

- Pour en savoir plus sur l'utilisation de NVMe/FC avec vSphere 7.0, consultez cette "Guide de configuration d'hôte NVMe/FC de ONTAP" et "TR-4684"La figure suivante décrit la connectivité multivoie d'un hôte vSphere vers un LUN ONTAP.



NFS

ONTAP est, entre autres, une baie NAS scale-out de grande qualité. ONTAP permet à VMware vSphere d'accéder simultanément aux datastores connectés par NFS à partir de nombreux hôtes VMware ESXi, ce qui dépasse de loin les limites imposées aux systèmes de fichiers VMFS. L'utilisation de NFS avec vSphere offre des avantages en termes de facilité d'utilisation et d'efficacité du stockage, comme indiqué dans la "les datastores" section.

Nous vous recommandons les meilleures pratiques suivantes lorsque vous utilisez ONTAP NFS avec vSphere :

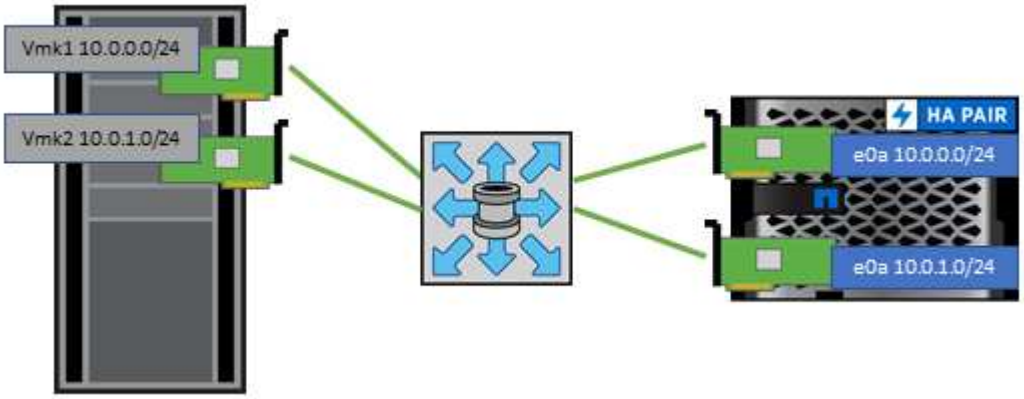
- Utilisez les outils ONTAP pour VMware vSphere (meilleure pratique la plus importante) :
  - Utilisez les outils ONTAP pour VMware vSphere pour provisionner les datastores, car ils simplifient automatiquement la gestion des règles d'exportation.
  - Lors de la création de datastores pour clusters VMware avec le plug-in, sélectionnez le cluster plutôt qu'un seul serveur ESX. Ce choix permet de monter automatiquement le datastore sur tous les hôtes du cluster.
  - Utilisez la fonction de montage du plug-in pour appliquer les datastores existants aux nouveaux serveurs.
  - Lorsque vous n'utilisez pas les outils ONTAP pour VMware vSphere, utilisez une export policy unique pour tous les serveurs ou pour chaque cluster de serveurs où un contrôle d'accès supplémentaire est nécessaire.
- Utiliser une interface logique (LIF) unique pour chaque SVM sur chaque nœud du cluster ONTAP. Les recommandations précédentes d'une LIF par datastore ne sont plus nécessaires. L'accès direct (LIF et datastore sur le même nœud) est préférable, mais ne vous inquiétez pas pour l'accès indirect, car l'effet de performance est généralement minimal (microsecondes).
- Si vous utilisez fpolicy, veillez à exclure les fichiers .lck car ils sont utilisés par vSphere pour le verrouillage à chaque mise sous tension d'une machine virtuelle.

- Toutes les versions de VMware vSphere actuellement prises en charge peuvent utiliser NFS v3 et v4.1. Le support officiel de nconnect a été ajouté à vSphere 8.0 mise à jour 2 pour NFS v3 et mise à jour 3 pour NFS v4.1. Pour NFS v4.1, vSphere continue à prendre en charge l'agrégation de sessions, l'authentification Kerberos et l'authentification Kerberos avec intégrité. Il est important de noter que l'agrégation de session nécessite ONTAP 9.14.1 ou une version ultérieure. Vous pouvez en savoir plus sur la fonction nconnect et sur la manière dont elle améliore les performances à "Fonctionnalité NFSv3 nconnect avec NetApp et VMware".

> - La valeur maximale de nconnect dans vSphere 8 est 4 et la valeur par défaut est 1. La limite de valeur maximale dans vSphere peut être augmentée par hôte grâce à des paramètres avancés, mais elle n'est généralement pas nécessaire.
> - Une valeur de 4 est recommandée pour les environnements nécessitant des performances supérieures à celles d'une seule connexion TCP.
> - Sachez que ESXi a une limite de 256 connexions NFS et que chaque connexion nconnect compte pour ce total. Par exemple, deux datastores avec nconnect=4 compteraient comme huit connexions au total.
> - Il est important de tester l'impact de nconnect sur les performances de votre environnement avant d'implémenter des modifications à grande échelle dans les environnements de production.

- Notez que NFS v3 et NFS v4.1 utilisent différents mécanismes de verrouillage. NFS v3 utilise un verrouillage côté client, tandis que NFS v4.1 utilise un verrouillage côté serveur. Bien qu'un volume ONTAP puisse être exporté via les deux protocoles, ESXi ne peut monter qu'un datastore via un protocole. Cependant, cela ne signifie pas que d'autres hôtes ESXi ne peuvent pas monter le même datastore via une version différente. Pour éviter tout problème, il est essentiel de spécifier la version du protocole à utiliser lors du montage, en veillant à ce que tous les hôtes utilisent la même version et, par conséquent, le même style de verrouillage. Il est essentiel d'éviter de mélanger les versions NFS entre les hôtes. Si possible, utilisez les profils hôtes pour vérifier la conformité.
  - Étant donné qu'il n'existe pas de conversion automatique de datastore entre NFS v3 et NFS v4.1, créez un nouveau datastore NFSv4.1 et utilisez Storage vMotion pour migrer les machines virtuelles vers le nouveau datastore.
  - Pour connaître les niveaux de correctifs ESXi requis pour la prise en charge, reportez-vous aux notes du tableau d'interopérabilité NFS v4.1 dans le"Matrice d'interopérabilité NetApp".
- Comme indiqué à la "paramètres", si vous n'utilisez pas vSphere CSI pour Kubernetes, vous devez définir newSyncInterval par "VMware KB 386364"
- Les règles d'export NFS permettent de contrôler l'accès des hôtes vSphere. Vous pouvez utiliser une seule règle avec plusieurs volumes (datastores). Avec NFS, ESXi utilise le style de sécurité sys (UNIX) et requiert l'option de montage racine pour exécuter les VM. Dans ONTAP, cette option est appelée superutilisateur et, lorsque l'option superutilisateur est utilisée, il n'est pas nécessaire de spécifier l'ID utilisateur anonyme. Notez que les règles d'export-policy avec des valeurs différentes pour `-anon` et `-allow-suid` peuvent causer des problèmes de découverte de SVM avec les outils ONTAP. Les adresses IP doivent être séparées par des virgules, sans espaces dans les adresses de port vmkernel qui montés dans les datastores. Voici un exemple de règle de stratégie :
  - Protocole d'accès : nfs (qui inclut nfs3 et nfs4)
  - Liste des noms d'hôte, adresses IP, groupes réseau ou domaines correspondant au client : 192.168.42.21,192.168.42.22
  - Règle d'accès RO : tous
  - Règle d'accès RW : tous

    - ID utilisateur auquel les utilisateurs anonymes sont mappés : 65534
    - Types de sécurité superutilisateur : tous
    - Honorez les bits setuid dans SETATTR : TRUE
    - Autoriser la création de périphériques : vrai
- Si le plug-in NetApp NFS pour VMware VAAI est utilisé, le protocole doit être défini comme `nfs` lors de la création ou de la modification de la règle d'export policy. Le protocole NFSv4 est requis pour que le déchargement des copies VAAI fonctionne, et la spécification du protocole comme `nfs` inclut automatiquement les versions NFSv3 et NFSv4. Cette opération est requise même si le type de datastore est créé en tant que NFS v3.
- Les volumes des datastores NFS sont rassemblés dans le volume racine du SVM. Par conséquent, ESXi doit également avoir accès au volume racine pour naviguer et monter des volumes de datastores. La export policy pour le volume root, et pour tout autre volume dans lequel la jonction du volume de datastore est imbriquée, doit inclure une règle ou des règles pour les serveurs ESXi leur accordant un accès en lecture seule. Voici un exemple de règle pour le volume racine, également à l'aide du plug-in VAAI :
    - Protocole d'accès : nfs
    - Comparaison avec le client : 192.168.42.21,192.168.42.22
    - Règle d'accès RO : sys
    - Règle d'accès RW : jamais (meilleure sécurité pour le volume racine)
    - UID anonyme
    - Superutilisateur : sys (également requis pour le volume racine avec VAAI)
- Bien que ONTAP offre une structure d'espace de noms de volume flexible permettant d'organiser les volumes dans une arborescence à l'aide de jonctions, cette approche n'a aucune valeur pour vSphere. Il crée un répertoire pour chaque machine virtuelle à la racine du datastore, quelle que soit la hiérarchie de l'espace de noms du stockage. Il est donc recommandé de simplement monter le Junction path pour les volumes pour vSphere au volume root du SVM, c'est-à-dire comment les outils ONTAP pour VMware vSphere provisionne les datastores. Sans chemins de jonction imbriqués, aucun volume ne dépend d'aucun volume autre que le volume root et que mettre un volume hors ligne ou le détruire, même intentionnellement, n'affecte pas le chemin d'accès aux autres volumes.
- Une taille de bloc de 4 Ko convient parfaitement aux partitions NTFS sur les datastores NFS. La figure suivante décrit la connectivité d'un hôte vSphere vers un datastore NFS ONTAP.



Le tableau suivant répertorie les versions NFS et les fonctionnalités prises en charge.

Fonctionnalités de vSphere	NFSv3	NFSv4.1
VMotion et Storage vMotion	Oui.	Oui.
Haute disponibilité	Oui.	Oui.
Tolérance aux pannes	Oui.	Oui.
DRS	Oui.	Oui.
Profils hôtes	Oui.	Oui.
DRS de stockage	Oui.	Non
Contrôle des E/S du stockage	Oui.	Non
SRM	Oui.	Non
Volumes virtuels	Oui.	Non
Accélération matérielle (VAAI)	Oui.	Oui.
Authentification Kerberos	Non	Oui (optimisé avec vSphere 6.5 et versions ultérieures pour prendre en charge AES et krb5i)
Prise en charge des chemins d'accès	Non	Oui (ONTAP 9.14.1)

Volumes FlexGroup

Utilisez des volumes ONTAP et FlexGroup avec VMware vSphere pour disposer de datastores simples et évolutifs exploitant toute la puissance d'un cluster ONTAP.

ONTAP 9.8, ainsi que les outils ONTAP pour VMware vSphere 9.8-9.13 et le plug-in SnapCenter pour VMware 4.4 et les versions ultérieures, ont ajouté la prise en charge des datastores FlexGroup avec volumes dans vSphere. Les volumes FlexGroup simplifient la création de grands datastores et créent automatiquement les volumes distribués nécessaires sur le cluster ONTAP afin d'optimiser les performances d'un système ONTAP.

Utilisez les volumes FlexGroup avec vSphere si vous avez besoin d'un datastore vSphere unique et évolutif doté de la puissance d'un cluster ONTAP complet, ou si vous disposez de charges de travail de clonage très volumineuses qui peuvent tirer parti du mécanisme de clonage FlexGroup en conservant constamment le cache de clone au chaud.

Copie auxiliaire

Outre les tests approfondis du système avec les charges de travail vSphere, ONTAP 9.8 a ajouté un nouveau mécanisme de déchargement des copies pour les datastores FlexGroup. Ce nouveau système utilise un moteur de copie amélioré pour répliquer les fichiers entre les composants en arrière-plan tout en permettant l'accès à la source et à la destination. Ce cache local constitutif est ensuite utilisé pour instancier rapidement les clones de machine virtuelle à la demande.

Pour activer le déchargement de copie optimisé pour FlexGroup, reportez-vous à la section "Comment configurer les volumes ONTAP FlexGroup pour permettre la copie auxiliaire VAAI"

Si vous utilisez le clonage VAAI, mais que le clonage n'est pas suffisant pour maintenir le cache chaud, vos clones ne seront peut-être pas plus rapides qu'une copie basée sur hôte. Si c'est le cas, vous pouvez régler le délai d'expiration du cache pour mieux répondre à vos besoins.

Prenons le scénario suivant :

- Vous avez créé un nouveau FlexGroup avec 8 composants
- Le délai d'expiration du cache pour le nouveau FlexGroup est défini sur 160 minutes

Dans ce scénario, les 8 premiers clones à terminer seront des copies complètes, et non des clones de fichiers locaux. Tout clonage supplémentaire de cette machine virtuelle avant l'expiration du délai de 160 secondes utilisera le moteur de clonage de fichiers à l'intérieur de chaque composant de manière circulaire pour créer des copies quasi immédiates réparties uniformément sur les volumes constitutifs.

Chaque nouvelle tâche de clonage reçue par un volume réinitialise le délai d'expiration. Si un volume composant de l'exemple FlexGroup ne reçoit pas de requête de clone avant le délai d'expiration, le cache de cette machine virtuelle sera effacé et le volume devra être à nouveau rempli. De même, si la source du clone d'origine change (par exemple, si vous avez mis à jour le modèle), le cache local de chaque composant sera invalidé pour éviter tout conflit. Comme indiqué précédemment, le cache peut être réglé en fonction des besoins de votre environnement.

Pour plus d'informations sur l'utilisation des volumes FlexGroup avec VAAI, consultez l'article de la base de connaissances suivant : "VAAI : comment la mise en cache fonctionne-t-elle avec les volumes FlexGroup ?"

Dans les environnements où vous ne pouvez pas tirer pleinement parti du cache FlexGroup, mais où vous avez toujours besoin d'un clonage rapide entre plusieurs volumes, envisagez d'utiliser les vVols. Le clonage entre volumes avec vVols est beaucoup plus rapide qu'avec les datastores traditionnels et ne repose pas sur un cache.

Paramètres QoS

La configuration de la qualité de service au niveau FlexGroup à l'aide de ONTAP System Manager ou du shell du cluster est prise en charge, mais elle ne prend pas en charge la reconnaissance des machines virtuelles ni l'intégration de vCenter.

La qualité de service (IOPS max/min) peut être définie sur des VM individuelles ou sur toutes les VM d'un datastore à ce moment dans l'interface utilisateur vCenter ou via les API REST à l'aide des outils ONTAP. La définition de la qualité de service sur toutes les VM remplace tous les paramètres distincts par VM. Les paramètres ne s'étendent pas ultérieurement aux nouvelles machines virtuelles ou aux machines virtuelles migrées ; définissez la qualité de service sur les nouvelles machines virtuelles ou appliquez à nouveau la qualité de service à toutes les machines virtuelles du datastore.

Notez que VMware vSphere traite toutes les E/S d'un datastore NFS comme une seule file d'attente par hôte, et que la limitation de la qualité de service sur une machine virtuelle peut avoir un impact sur les performances des autres machines virtuelles du même datastore pour cet hôte. Cela contraste avec les vVols qui peuvent maintenir leurs paramètres de politique de QoS s'ils migrent vers un autre datastore et n'ont pas d'impact sur les E/S d'autres machines virtuelles lorsqu'ils sont restreints.

Métriques

ONTAP 9.8 a également ajouté de nouveaux metrics de performance basés sur des fichiers (IOPS, débit et latence) pour FlexGroup Files. Ces metrics peuvent être consultées dans les outils ONTAP pour les rapports sur les machines virtuelles et le tableau de bord VMware vSphere. Les outils ONTAP pour le plug-in VMware vSphere vous permettent également de définir des règles de qualité de service (QoS) en combinant des IOPS minimales et/ou maximales. Ils peuvent être définis au sein de toutes les machines virtuelles d'un datastore ou individuellement pour des machines virtuelles spécifiques.

Et des meilleures pratiques

- Utilisez les outils ONTAP pour créer des datastores FlexGroup afin de vous assurer que votre FlexGroup est créé de manière optimale et que les règles d'exportation sont configurées pour correspondre à votre environnement vSphere. Cependant, après avoir créé le volume FlexGroup avec les outils ONTAP, vous constaterez que tous les nœuds de votre cluster vSphere utilisent une seule adresse IP pour monter le datastore. Cela pourrait entraîner un goulot d'étranglement sur le port réseau. Pour éviter ce problème, démontez le datastore, puis remontez-le à l'aide de l'assistant standard vSphere datastore en utilisant un nom DNS round-Robin qui équilibre la charge entre les LIF du SVM. Après le remontage, les outils ONTAP pourront à nouveau gérer le datastore. Si les outils ONTAP ne sont pas disponibles, utilisez les paramètres par défaut de FlexGroup et créez votre règle d'export en suivant les instructions de la section "Datastores et protocoles - NFS".
- Lors du dimensionnement d'un datastore FlexGroup, n'oubliez pas que le FlexGroup est constitué de plusieurs petits volumes FlexVol qui créent un espace de noms plus important. Par conséquent, dimensionnez le datastore pour qu'il soit au moins 8 fois (en supposant que les 8 composants par défaut) la taille de votre fichier VMDK le plus volumineux, plus une marge inutilisée de 10 à 20 % pour permettre un rééquilibrage flexible. Par exemple, si votre environnement comporte 6 To de VMDK, dimensionnez le datastore FlexGroup d'une capacité inférieure à 52,8 To (6 x 8 + 10 %).
- VMware et NetApp prennent en charge la mise en circuit de session NFSv4.1 à partir de ONTAP 9.14.1. Pour plus d'informations sur les versions, consultez les notes relatives à la matrice d'interopérabilité NetApp NFS 4.1 (IMT). NFSv3 ne prend pas en charge plusieurs chemins physiques vers un volume, mais prend en charge nconnect à partir de vSphere 8.0U2. Pour plus d'informations sur nconnect, consultez le "Fonctionnalité NFSv3 nConnect avec NetApp et VMware".
- Utilisez le plug-in NFS pour VMware VAAI pour la copie auxiliaire. Notez que même si le clonage est amélioré dans un datastore FlexGroup, comme mentionné précédemment, ONTAP n'offre pas d'avantages significatifs en termes de performances par rapport à la copie hôte ESXi lors de la copie de machines virtuelles entre des volumes FlexVol et/ou FlexGroup. Prenez donc en compte vos charges de travail de clonage lorsque vous décidez d'utiliser des volumes VAAI ou FlexGroup. L'une des façons d'optimiser le clonage basé sur FlexGroup consiste à modifier le nombre de volumes constitutifs. Tout comme le réglage du délai d'expiration du cache mentionné précédemment.
- Utilisez les outils ONTAP pour VMware vSphere 9.8-9.13 pour surveiller les performances des machines virtuelles FlexGroup à l'aide de metrics ONTAP (tableaux de bord et rapports sur les machines virtuelles) et gérer la qualité de service sur chaque machine virtuelle. Ces metrics ne sont pas encore disponibles via les commandes ou les API ONTAP.
- Le plug-in SnapCenter pour VMware vSphere version 4.4 et ultérieure prend en charge la sauvegarde et la restauration des machines virtuelles dans un datastore FlexGroup sur le système de stockage principal. Le distributeur sélectif 4.6 ajoute la prise en charge de SnapMirror pour les datastores basés sur FlexGroup. L'utilisation de snapshots basés sur les baies et de la réplication est le moyen le plus efficace de protéger vos données.

Configuration du réseau

La configuration des paramètres réseau lors de l'utilisation de vSphere avec des systèmes exécutant ONTAP est simple et similaire à celle des autres configurations réseau.

Voici quelques points à prendre en compte :

- Trafic du réseau de stockage séparé des autres réseaux Un réseau distinct peut être obtenu à l'aide d'un VLAN dédié ou de commutateurs distincts pour le stockage. Si le réseau de stockage partage des chemins physiques, tels que des liaisons ascendantes, vous pouvez avoir besoin de la qualité de service ou de ports supplémentaires pour garantir une bande passante suffisante. Ne connectez pas les hôtes

directement au stockage, sauf si votre guide de solution le demande expressément ; utilisez les commutateurs pour disposer de chemins redondants et permettez à VMware HA de fonctionner sans intervention.

- Les trames Jumbo doivent être utilisées si elles sont prises en charge par votre réseau. Si elles sont utilisées, assurez-vous qu'elles sont configurées de manière identique sur tous les périphériques réseau, VLAN, etc. Dans le chemin entre le stockage et l'hôte ESXi. Vous pourriez voir des problèmes de performances ou de connexion. La MTU doit également être définie de manière identique sur le switch virtuel ESXi, le port VMkernel et également sur les ports physiques ou les groupes d'interface de chaque nœud ONTAP.
- NetApp recommande uniquement de désactiver le contrôle de flux réseau sur les ports d'interconnexion de cluster au sein d'un cluster ONTAP. NetApp ne formule aucune autre recommandation concernant les meilleures pratiques en matière de contrôle de flux pour les ports réseau restants utilisés pour le trafic de données. Vous devez l'activer ou la désactiver si nécessaire. Voir "TR-4182" pour plus d'informations sur le contrôle de flux.
- Lorsque les baies de stockage ESXi et ONTAP sont connectées aux réseaux de stockage Ethernet, NetApp recommande de configurer les ports Ethernet auxquels ces systèmes se connectent en tant que ports de périphérie RSTP (Rapid Spanning Tree Protocol) ou en utilisant la fonctionnalité Cisco PortFast. NetApp recommande d'activer la fonction de jonction Spanning-Tree PortFast dans les environnements qui utilisent la fonction Cisco PortFast et dont le agrégation VLAN 802.1Q est activée soit au serveur ESXi, soit aux baies de stockage ONTAP.
- NetApp recommande les meilleures pratiques suivantes pour l'agrégation de liens :
  - Utilisez des commutateurs qui prennent en charge l'agrégation de liens des ports sur deux châssis de commutateurs distincts grâce à une approche de groupe d'agrégation de liens multichâssis, telle que Virtual PortChannel (VPC) de Cisco.
  - Désactiver LACP pour les ports de switch connectés à ESXi, sauf si vous utilisez dvswitches 5.1 ou version ultérieure avec LACP configuré.
  - Utilisez LACP pour créer des agrégats de liens pour les systèmes de stockage ONTAP avec des groupes d'interface multimode dynamiques avec un hachage IP.
  - Utilisez une stratégie de regroupement de hachage IP sur ESXi.

Le tableau suivant fournit un récapitulatif des éléments de configuration réseau et indique l'emplacement d'application des paramètres.

Élément	VMware ESXi	Commutateur	Nœud	SVM
Adresse IP	VMkernel	Non**	Non**	Oui.
Agrégation de liens	Commutateur virtuel	Oui.	Oui.	Non*
VLAN	Groupes de ports VMKernel et VM	Oui.	Oui.	Non*
Contrôle de flux	NIC	Oui.	Oui.	Non*
Spanning Tree	Non	Oui.	Non	Non
MTU (pour les trames jumbo)	Commutateur virtuel et port VMkernel (9000)	Oui (défini sur max)	Oui (9000)	Non*
Groupes de basculement	Non	Non	Oui (créer)	Oui (sélectionner)

*Les LIF SVM se connectent aux ports, aux groupes d'interface ou aux interfaces VLAN dotés de VLAN, MTU et d'autres paramètres. Cependant, les paramètres ne sont pas gérés au niveau de la SVM.

**Ces périphériques ont leur propre adresse IP pour la gestion, mais ces adresses ne sont pas utilisées dans le contexte du réseau de stockage VMware ESXi.

SAN (FC, NVMe/FC, iSCSI, NVMe/TCP), RDM

ONTAP propose un stockage bloc haute performance pour VMware vSphere à l'aide d'iSCSI classiques et du protocole FCP (Fibre Channel Protocol). Il prend également en charge NVMe/FC et NVMe/TCP, le protocole bloc nouvelle génération hautement efficace et performant, NVMe over Fabrics (NVMe-of).

Pour connaître les meilleures pratiques détaillées en matière d'implémentation de protocoles par blocs pour le stockage de machines virtuelles avec vSphere et ONTAP, reportez-vous à la section "Datastores et protocoles - SAN"

NFS

VSphere permet aux clients d'utiliser des baies NFS de classe entreprise pour fournir un accès simultané aux datastores à tous les nœuds d'un cluster ESXi. Comme mentionné dans cette "les datastores" section, NFS avec vSphere offre des avantages en termes de visibilité sur la facilité d'utilisation et l'efficacité du stockage.

Pour connaître les meilleures pratiques recommandées, reportez-vous à la section "Datastores et protocoles - NFS"

Connexion directe au réseau

Les administrateurs du stockage préfèrent parfois simplifier leurs infrastructures en supprimant les commutateurs réseau de la configuration. Cela peut être pris en charge dans certains scénarios. Cependant, il y a quelques limitations et mises en garde à prendre en compte.

ISCSI et NVMe/TCP

Un hôte utilisant iSCSI ou NVMe/TCP peut être directement connecté à un système de stockage et fonctionner normalement. La raison en est le chemin d'accès. Les connexions directes à deux contrôleurs de stockage distincts donnent lieu à deux chemins de flux de données indépendants. La perte du chemin, du port ou du contrôleur n'empêche pas l'autre chemin d'être utilisé.

NFS

Vous pouvez utiliser un stockage NFS à connexion directe, mais avec une limitation importante : le basculement ne fonctionnera pas sans script important, ce qui incombera au client.

Ce qui complique la reprise après incident avec un stockage NFS à connexion directe, c'est le routage qui se produit sur le système d'exploitation local. Par exemple, supposons qu'un hôte a une adresse IP 192.168.1.1/24 et qu'il est directement connecté à un contrôleur ONTAP avec une adresse IP 192.168.1.50/24. Lors du basculement, cette adresse 192.168.1.50 peut basculer vers l'autre contrôleur et sera disponible pour l'hôte, mais comment l'hôte peut-il détecter sa présence ? L'adresse 192.168.1.1 d'origine existe toujours sur la carte réseau hôte qui ne se connecte plus à un système opérationnel. Le trafic destiné à 192.168.1.50 continuerait d'être envoyé à un port réseau inutilisable.

Le second NIC du système d'exploitation peut être configuré sur 19 2.168.1.2 et serait capable de communiquer avec l'adresse en panne sur 192.168.1.50, mais les tables de routage locales auraient par défaut l'utilisation d'une adresse **et d'une seule adresse** pour communiquer avec le sous-réseau 192.168.1.0/24. Un administrateur système pourrait créer un framework de scripts qui détecterait une

connexion réseau défaillante et modifierait les tables de routage locales ou rendrait les interfaces « up and down ». La procédure exacte dépend du système d'exploitation utilisé.

Dans la pratique, les clients NetApp disposent d'un protocole NFS à connexion directe, mais généralement uniquement pour les charges de travail où une pause des E/S est acceptable pendant les basculements. Lorsque des montages durs sont utilisés, aucune erreur d'E/S ne doit se produire lors de ces pauses. L'E/S doit se bloquer jusqu'à ce que les services soient restaurés, soit par une restauration automatique, soit par une intervention manuelle pour déplacer les adresses IP entre les cartes réseau de l'hôte.

Connexion directe FC

Il n'est pas possible de connecter directement un hôte à un système de stockage ONTAP à l'aide du protocole FC. La raison en est l'utilisation de NPIV. Le WWN qui identifie un port FC ONTAP sur le réseau FC utilise un type de virtualisation appelé NPIV. Tout périphérique connecté à un système ONTAP doit pouvoir reconnaître un WWN NPIV. Aucun fournisseur actuel de HBA ne propose de HBA pouvant être installé sur un hôte et capable de prendre en charge une cible NPIV.

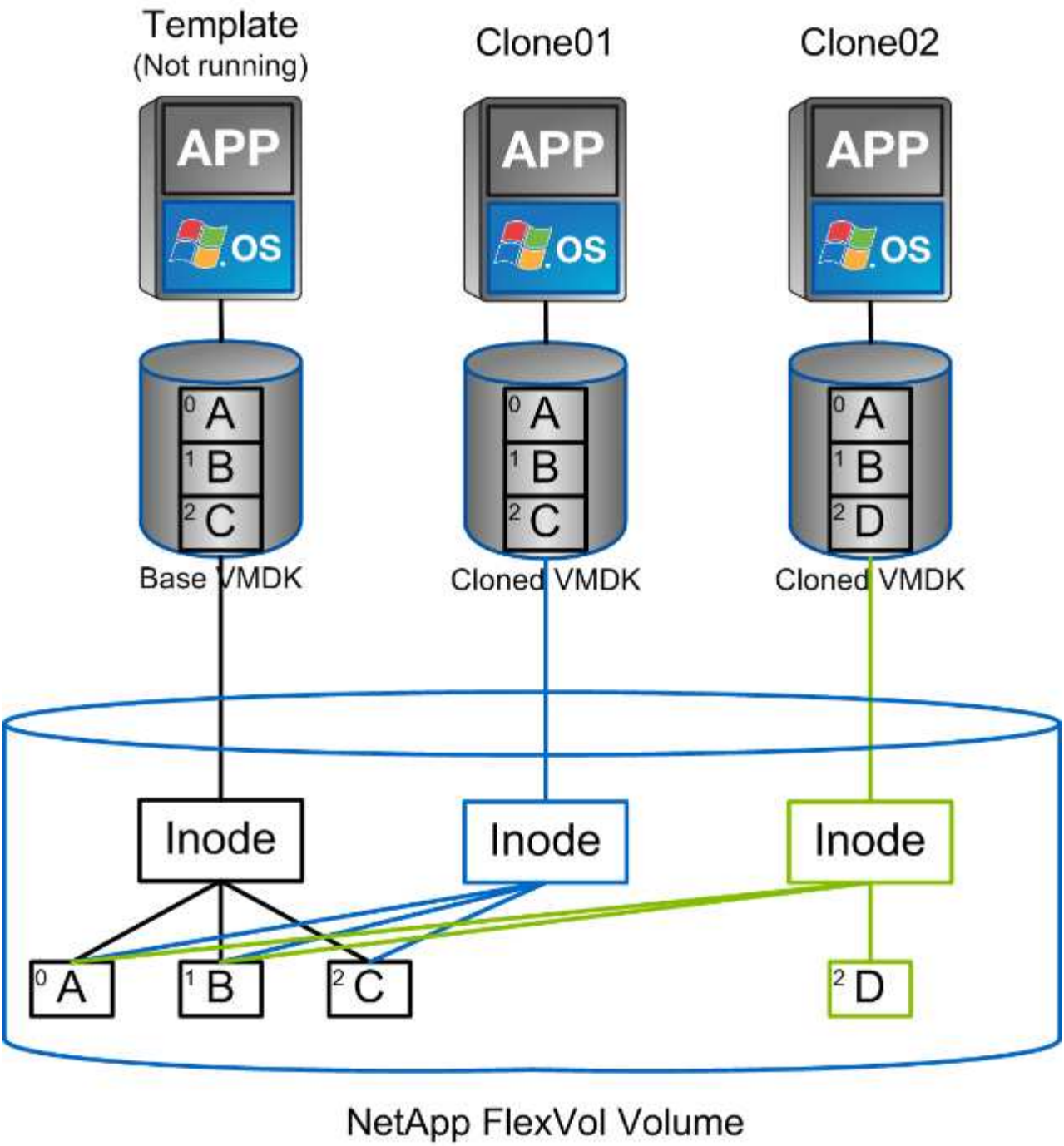
Clonage des VM et des datastores

Le clonage d'un objet de stockage vous permet de créer rapidement des copies pour ensuite les utiliser, par exemple le provisionnement de machines virtuelles supplémentaires, les opérations de sauvegarde/restauration, etc.

Dans vSphere, vous pouvez cloner une machine virtuelle, un disque virtuel, un volume virtuel ou un datastore. Une fois cloné, l'objet peut être davantage personnalisé, souvent par le biais d'un processus automatisé. VSphere prend en charge les clones de copie complète ainsi que les clones liés, pour assurer le suivi séparé des modifications apportées à l'objet d'origine.

Les clones liés permettent un gain d'espace considérable, mais ils augmentent la quantité d'E/S que vSphere gère pour la machine virtuelle, ce qui affecte les performances de cette machine virtuelle, et peut-être de l'hôte dans son ensemble. C'est pourquoi les clients NetApp utilisent souvent des clones basés sur des systèmes de stockage pour profiter d'un double avantage : une utilisation efficace du stockage et des performances supérieures.

La figure suivante représente le clonage ONTAP.



Le clonage peut être déchargé sur les systèmes qui exécutent ONTAP par le biais de plusieurs mécanismes, généralement au niveau des VM, vVol ou datastore. Ces champs d'application incluent :

- Vvols avec le fournisseur NetApp vSphere APIs for Storage Awareness (VASA). Les clones ONTAP sont utilisés pour prendre en charge les snapshots vVol gérés par vCenter. Ces snapshots sont peu encombrants avec un impact E/S minimal en termes de création et de suppression. Les machines virtuelles peuvent également être clonées via vCenter, qui sont également déchargées vers ONTAP, que ce soit dans un datastore/volume unique ou entre les datastores/volumes.
- Clonage et migration de vSphere à l'aide des API vSphere – intégration de baies (VAAI). Les opérations de clonage de VM peuvent être déchargées vers ONTAP dans les environnements SAN et NAS (NetApp fournit un plug-in ESXi pour VAAI for NFS). VSphere ne décharge les opérations que sur les machines virtuelles inactives (hors tension) d'un datastore NAS, tandis que les opérations sur les machines virtuelles actives (clonage et stockage vMotion) sont également déchargées pour SAN. ONTAP utilise l'approche la

plus efficace, basée sur la source et la destination. Cette fonctionnalité est également utilisée par "Vue horizon Omnissa".

- SRA (utilisée avec VMware Live site Recovery Manager/site Recovery Manager). Ici, des clones sont utilisés pour tester la restauration de la réplique de reprise après incident sans interruption.
- Sauvegarde et restauration à l'aide d'outils NetApp tels que SnapCenter. Les clones des machines virtuelles sont utilisés pour vérifier les opérations de sauvegarde et monter une sauvegarde de machines virtuelles afin de restaurer les fichiers individuels.

Le clonage ONTAP Offloaded peut être appelé par les outils VMware, NetApp et tiers. Les clones déchargés sur ONTAP présentent plusieurs avantages. Elles sont peu gourmandes en espace dans la plupart des cas, et n'ont besoin que de systèmes de stockage pour modifier les objets. Cela n'a aucun impact supplémentaire sur les performances en lecture et en écriture. Dans certains cas, le partage des blocs dans des caches haute vitesse améliore les performances. Ils délester également le serveur ESXi de la charge des cycles CPU et des E/S réseau. La fonctionnalité de déchargement des copies dans un datastore classique utilisant un FlexVol volume peut être rapide et efficace avec une licence FlexClone (incluse dans la licence ONTAP One), mais les copies entre des volumes FlexVol peuvent être plus lentes. Si vous maintenez les modèles de machine virtuelle comme source de clones, envisagez de les placer dans le volume du datastore (utilisez les dossiers ou les bibliothèques de contenu pour les organiser) afin de créer des clones rapides et compacts.

Vous pouvez également cloner un volume ou une LUN directement au sein de ONTAP afin de cloner un datastore. Grâce aux datastores NFS, la technologie FlexClone peut cloner un volume entier. Le clone peut être exporté depuis ONTAP et monté par ESXi en tant qu'autre datastore. Pour les datastores VMFS, ONTAP peut cloner une LUN au sein d'un volume ou d'un volume complet, y compris une ou plusieurs LUN au sein de celle-ci. Une LUN contenant un VMFS doit être mappée sur un groupe d'initiateurs ESXi, puis une nouvelle signature définie par ESXi doit être montée et utilisée comme datastore standard. Pour certains cas d'utilisation temporaire, un VMFS cloné peut être monté sans nouvelle signature. Une fois le datastore cloné, les ordinateurs virtuels internes peuvent être enregistrés, reconfigurés et personnalisés comme s'ils étaient individuellement clonés.

Dans certains cas, des fonctionnalités supplémentaires sous licence peuvent être utilisées pour améliorer le clonage, telles que SnapRestore pour la sauvegarde ou FlexClone. Ces licences sont souvent incluses dans les packs de licence sans frais supplémentaires. Une licence FlexClone est requise pour les opérations de clonage vVol, ainsi que pour la prise en charge des snapshots gérés d'un VVol (qui sont déchargés de l'hyperviseur vers ONTAP). Une licence FlexClone peut également améliorer certains clones VAAI lorsqu'ils sont utilisés dans un datastore/volume (création de copies instantanées et compactes à la place de copies de bloc). Elle est également utilisée par SRA pour tester la restauration d'une réplique de reprise après incident et SnapCenter pour les opérations de clonage, et pour parcourir les copies de sauvegarde afin de restaurer des fichiers individuels.

Protection des données

La sauvegarde et la restauration rapide de vos machines virtuelles sont les principaux avantages de ONTAP pour vSphere. Cette fonctionnalité peut être facilement gérée dans vCenter via le plug-in SnapCenter pour VMware vSphere. De nombreux clients améliorent leurs solutions de sauvegarde tierces avec SnapCenter pour exploiter la technologie Snapshot de ONTAP, car il offre le moyen le plus rapide et le plus simple de restaurer une machine virtuelle avec ONTAP. SnapCenter est disponible gratuitement pour les clients disposant d'une licence ONTAP One, et d'autres packs de licences peuvent également être disponibles.

De plus, le plug-in SnapCenter pour VMware peut s'intégrer avec "BlueXP pour la sauvegarde et la restauration des machines virtuelles", offrant ainsi des solutions de sauvegarde efficaces 3 à 2-1 pour la

plupart des systèmes ONTAP. Notez que des frais peuvent s'appliquer en cas d'utilisation de la sauvegarde et de la restauration BlueXP pour des machines virtuelles avec des services premium, tels que les magasins d'objets pour un stockage de sauvegarde supplémentaire. Cette section décrit les différentes options disponibles pour protéger vos machines virtuelles et vos datastores.

Snapshots de volumes NetApp ONTAP

Utilisez les snapshots pour créer des copies rapides de votre machine virtuelle ou de votre datastore sans affecter les performances, puis envoyez-les à un système secondaire à l'aide de SnapMirror pour une protection des données hors site à plus long terme. Cette approche réduit l'espace de stockage et la bande passante réseau en stockant uniquement les informations modifiées.

Les snapshots sont une fonctionnalité clé de ONTAP, ce qui vous permet de créer des copies instantanées de vos données. Ils sont compacts et peuvent être créés rapidement, ce qui en fait la solution idéale pour la protection des machines virtuelles et des datastores. Les copies Snapshot peuvent être utilisées à diverses fins, notamment la sauvegarde, la restauration et le test. Ces snapshots sont différents des snapshots VMware (cohérence) et conviennent à une protection à long terme. Les snapshots gérés par vCenter de VMware sont uniquement recommandés pour une utilisation à court terme en raison des performances et d'autres effets. Voir "Limites des snapshots"pour plus de détails.

Les snapshots sont créés au niveau du volume et peuvent être utilisés pour protéger toutes les machines virtuelles et tous les datastores de ce volume. Cela signifie que vous pouvez créer un snapshot d'un datastore complet, qui inclut toutes les machines virtuelles de ce datastore.

Pour les datastores NFS, vous pouvez facilement afficher les fichiers des machines virtuelles dans les snapshots en parcourant le répertoire .snapshots. Cela vous permet d'accéder rapidement aux fichiers et de les restaurer à partir d'une copie Snapshot sans avoir à utiliser de solution de sauvegarde spécifique.

Pour les datastores VMFS, vous pouvez créer un FlexClone du datastore basé sur le snapshot souhaité. Cela vous permet de créer un nouveau datastore basé sur le snapshot, qui peut être utilisé à des fins de test ou de développement. La FlexClone ne consomme de l'espace pour les modifications effectuées qu'après la création de l'instantané, ce qui en fait un moyen compact de créer une copie du datastore. Une fois la FlexClone créée, vous pouvez mapper la LUN ou l'espace de noms vers un hôte ESXi comme un datastore standard. Cela vous permet non seulement de restaurer des fichiers de machine virtuelle spécifiques, mais également de créer rapidement des environnements de test ou de développement basés sur des données de production sans affecter les performances de l'environnement de production.

Pour plus d'informations sur les instantanés, reportez-vous à la documentation ONTAP. Les liens suivants fournissent des détails supplémentaires : "Copies Snapshot locales ONTAP" "Réplication ONTAP SnapMirror"

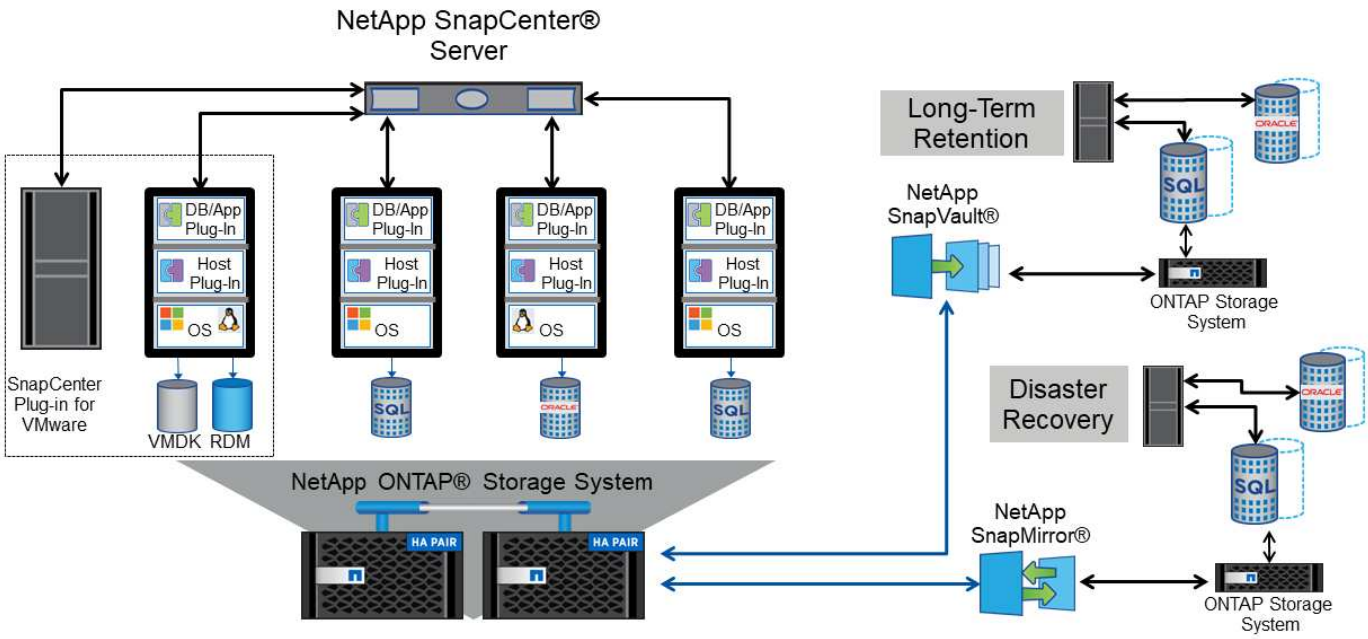
Plug-in SnapCenter pour VMware vSphere

SnapCenter vous permet de créer des règles de sauvegarde qui peuvent être appliquées à plusieurs tâches. Ces règles peuvent définir des fonctionnalités de planification, de conservation, de réplication et autres. Ils continuent d'autoriser une sélection facultative de snapshots cohérents avec les machines virtuelles, ce qui exploite la capacité de l'hyperviseur à suspendre les E/S avant de prendre un snapshot VMware. Cependant, en raison de l'impact des snapshots VMware sur les performances, ils ne sont généralement pas recommandés sauf si vous devez suspendre le système de fichiers invité. Utilisez plutôt les snapshots pour une protection générale et des outils applicatifs tels que les plug-ins d'applications SnapCenter pour protéger les données transactionnelles comme SQL Server ou Oracle.

Ces plug-ins offrent des fonctionnalités étendues pour protéger les bases de données dans les environnements physiques et virtuels. Avec vSphere, vous pouvez les utiliser pour protéger les bases de données SQL Server ou Oracle où les données sont stockées sur des LUN RDM, des vVols ou des namespaces NVMe/TCP et des LUN iSCSI directement connectées au système d'exploitation invité, ou des

fichiers VMDK sur des datastores VMFS ou NFS. Les plug-ins permettent de spécifier différents types de sauvegardes de bases de données, de prendre en charge la sauvegarde en ligne ou hors ligne et de protéger les fichiers de base de données ainsi que les fichiers journaux. Outre la sauvegarde et la restauration, les plug-ins prennent également en charge le clonage des bases de données à des fins de développement ou de test.

La figure suivante représente un exemple de déploiement SnapCenter.



Pour plus d'informations sur le dimensionnement, reportez-vous au "Guide de dimensionnement du plug-in SnapCenter pour VMware vSphere"

Outils ONTAP pour VMware vSphere avec VMware Live site Recovery

Les outils ONTAP pour VMware vSphere (OT4VS) sont un plug-in gratuit qui permet une intégration transparente entre VMware vSphere et NetApp ONTAP. Il vous permet de gérer votre stockage ONTAP directement à partir du client Web vSphere, ce qui facilite les tâches telles que le provisionnement du stockage, la gestion de la réplication et la surveillance des performances.

Pour améliorer les fonctionnalités de reprise après incident, envisagez d'utiliser NetApp SRA pour ONTAP, qui fait partie des outils ONTAP pour VMware vSphere, en plus de VMware Live site Recovery Manager (anciennement site Recovery Manager). Cet outil prend non seulement en charge la réplication des datastores sur un site de reprise d'activité à l'aide de SnapMirror, mais il permet également de réaliser des tests sans interruption dans l'environnement de reprise après incident en clonant les datastores répliqués. De plus, les fonctionnalités d'automatisation intégrées rationalisent la restauration en cas d'incident et reprotègent la production après la résolution d'une panne.

Reprise d'activité BlueXP

BlueXP Disaster Recovery (DR) est un service cloud qui apporte une solution complète pour protéger vos données et vos applications en cas d'incident. La solution offre de nombreuses fonctionnalités, notamment des fonctionnalités de basculement et de restauration automatisés, plusieurs points de restauration à un point dans le temps, une reprise d'activité cohérente avec les applications et la prise en charge des systèmes ONTAP sur site et dans le cloud. La reprise après incident BlueXP est conçue pour fonctionner de manière transparente avec ONTAP et votre environnement VMware vSphere, offrant ainsi une solution unifiée pour la reprise après

incident.

Cluster de stockage vSphere Metro (vMSC) avec NetApp MetroCluster et la synchronisation active SnapMirror

Enfin, pour une protection optimale des données, envisagez une configuration VMware vSphere Metro Storage Cluster (vMSC) utilisant NetApp MetroCluster. VMSC est une solution NetApp certifiée VMware qui utilise une réplication synchrone, offrant les mêmes avantages qu'un cluster haute disponibilité, mais distribuée sur des sites distincts, pour une protection contre les incidents sur site. La solution NetApp SnapMirror Active Sync, avec ASA et AFF, et MetroCluster avec AFF, offre des configurations économiques pour la réplication synchrone avec restauration transparente en cas de défaillance d'un composant de stockage unique, ainsi qu'une restauration transparente en cas de synchronisation active SnapMirror ou une restauration à commande unique en cas d'incident sur site avec MetroCluster. VMSC est décrit plus en détail dans "TR-4128".

La qualité de service (QoS)

Les limites de débit sont utiles pour contrôler les niveaux de service, gérer les charges de travail inconnues ou tester les applications avant le déploiement pour s'assurer qu'elles n'affectent pas les autres charges de travail en production. Elles peuvent également être utilisées pour contraindre une charge de travail dominante après son identification.

Prise en charge des règles de QoS de ONTAP

Les systèmes qui exécutent ONTAP peuvent utiliser la fonction QoS du stockage pour limiter le débit en Mbit/s et/ou en E/S par seconde (IOPS) de différents objets de stockage comme les fichiers, les LUN, les volumes ou des SVM entiers.

Des niveaux minimaux de service basés sur des IOPS sont également pris en charge pour assurer des performances prévisibles pour les objets SAN d'ONTAP 9.2 et pour les objets NAS d'ONTAP 9.3.

Le débit maximal de QoS sur un objet peut être défini en Mbit/s et/ou IOPS. Si les deux sont utilisés, la première limite atteinte est appliquée par ONTAP. Une charge de travail peut contenir plusieurs objets et une règle de QoS peut être appliquée à un ou plusieurs workloads. Lorsqu'une règle est appliquée à plusieurs workloads, celle-ci partage la limite totale de la règle. Les objets imbriqués ne sont pas pris en charge (par exemple, les fichiers d'un volume ne peuvent pas chacun avoir leur propre stratégie). La valeur minimale de qualité de service ne peut être définie que dans les IOPS.

Les outils suivants sont actuellement disponibles pour la gestion des règles de QoS de ONTAP et leur application aux objets :

- INTERFACE DE LIGNE DE COMMANDES DE ONTAP
- ONTAP System Manager
- OnCommand Workflow Automation
- Active IQ Unified Manager
- Kit d'outils NetApp PowerShell pour ONTAP
- Outils ONTAP pour VMware vSphere VASA Provider

Pour affecter une QoS à une LUN, y compris VMFS et RDM, le SVM ONTAP (affiché comme vServer), le chemin LUN et le numéro de série peuvent être obtenus du menu systèmes de stockage de la page d'accueil des outils ONTAP pour VMware vSphere. Sélectionner le système de stockage (SVM), puis objets associés >

SAN. Utilisez cette approche lors de la spécification de QoS à l'aide de l'un des outils ONTAP.

Reportez-vous à la section "Contrôle des performances et présentation de la gestion" pour en savoir plus.

Datastores NFS non vVols

Une règle de qualité de service ONTAP peut être appliquée au datastore entier ou aux fichiers VMDK individuels qu'il contient. Toutefois, il est important de comprendre que toutes les machines virtuelles d'un datastore NFS traditionnel (non vVols) partagent une file d'attente d'E/S commune à partir d'un hôte donné. Si une règle de qualité de service ONTAP limite un ordinateur virtuel, toutes les E/S de ce datastore semblent alors restreintes pour cet hôte.

Exemple:
* Vous configurez une limite QoS sur vm1.vmdk pour un volume monté en tant que datastore NFS traditionnel par l'hôte esxi-01.
* Le même hôte (esxi-01) utilise vm2.vmdk et se trouve sur le même volume.
* Si vm1.vmdk est étranglé, alors vm2.vmdk semble également être étranglé car il partage la même file d'attente d'E/S avec vm1.vmdk.

Cela ne s'applique pas aux vVols.

À partir de vSphere 6.5, vous pouvez gérer les limites granulaires au niveau des fichiers sur les datastores non vVols en utilisant la gestion basée sur des règles de stockage (SPBM) avec le contrôle des E/S de stockage (SIOC) v2.

Pour plus d'informations sur la gestion des performances avec les règles SIOC et SPBM, reportez-vous aux liens suivants.

"Règles basées sur l'hôte SPBM : SIOC v2"
"Gestion des ressources d'E/S de stockage avec vSphere"

Pour affecter une politique de QoS à un VMDK sur NFS, suivez les consignes suivantes :

- La politique doit être appliquée au `vmname-flat.vmdk` qui contient l'image réelle du disque virtuel, pas le `vmname.vmdk` (fichier de descripteur de disque virtuel) ou `vmname.vmx` (Fichier de descripteur de machine virtuelle).
- N'appliquez pas de règles aux autres fichiers VM tels que les fichiers d'échange virtuels (`vmname.vswp`).
- Lors de l'utilisation du client Web vSphere pour trouver des chemins de fichiers (datastore > fichiers), notez qu'il combine les informations de l' `- flat.vmdk` et `. vmdk` et montre simplement un fichier avec le nom du `. vmdk` mais la taille du `- flat.vmdk`. Autres `-flat` dans le nom du fichier pour obtenir le chemin correct.

Les datastores FlexGroup offrent des fonctionnalités QoS améliorées lors de l'utilisation des outils ONTAP pour VMware vSphere 9.8 et versions ultérieures. Vous pouvez facilement définir la qualité de service sur toutes les machines virtuelles d'un datastore ou sur des machines virtuelles spécifiques. Consultez la section FlexGroup de ce rapport pour plus d'informations. Notez que les limitations de QoS mentionnées précédemment s'appliquent toujours avec les datastores NFS classiques.

Datastores VMFS

Avec des LUN ONTAP, les règles de qualité de service peuvent être appliquées au volume FlexVol qui contient les LUN ou les LUN individuelles, mais pas aux fichiers VMDK individuels car ONTAP ne connaît pas le système de fichiers VMFS.

Datastores vVols

La qualité de service minimale et/ou maximale peut être facilement définie sur des machines virtuelles individuelles ou des VMDK sans affecter les autres machines virtuelles ou VMDK à l'aide de la gestion basée sur des règles de stockage et des vVols.

Lors de la création du profil de capacité de stockage pour le conteneur vVol, spécifiez une valeur IOPS max et/ou min sous la fonctionnalité de performance, puis indiquez ce SCP avec la stratégie de stockage de la VM. Utilisez cette règle lors de la création de la machine virtuelle ou appliquez-la à une machine virtuelle existante.

> VVols a besoin des outils ONTAP pour VMware vSphere qui fonctionnent comme le VASA Provider pour ONTAP. Reportez-vous à "VMware vSphere Virtual volumes (vVols) avec ONTAP" la pour connaître les bonnes pratiques vVols.

QoS ONTAP et SIOC VMware

La QoS ONTAP et le contrôle des E/S du stockage VMware vSphere sont des technologies complémentaires que les administrateurs de stockage et vSphere peuvent utiliser conjointement pour gérer les performances des VM vSphere hébergées sur les systèmes exécutant ONTAP. Chaque outil a ses propres forces, comme le montre le tableau suivant. En raison des différents champs d'application de VMware vCenter et de ONTAP, certains objets peuvent être vus et gérés par un système et non par l'autre.

Propriété	QoS de ONTAP	SIOC VMware
Lorsqu'il est actif	La règle est toujours active	Actif en cas de conflit (latence du datastore supérieure au seuil)
Type d'unités	IOPS, Mo/sec	IOPS, partages
Étendue vCenter ou des applications	Plusieurs environnements vCenter, d'autres hyperviseurs et applications	Un seul serveur vCenter
Définir la qualité de service sur la machine virtuelle ?	VMDK sur NFS uniquement	VMDK sur NFS ou VMFS
Définir la qualité de service sur la LUN (RDM) ?	Oui.	Non
Définir la QoS sur LUN (VMFS) ?	Oui.	Oui (le datastore peut être limité)
Définir la qualité de service sur le volume (datastore NFS) ?	Oui.	Oui (le datastore peut être limité)
Qualité de service définie sur un SVM (locataire) ?	Oui.	Non
Approche basée sur des règles ?	Oui. Elles peuvent être partagées par toutes les charges de travail dans la règle ou appliquées en totalité à chaque charge de travail dans la règle.	Oui, avec vSphere 6.5 et versions ultérieures.
Licence requise	Inclus avec ONTAP	Enterprise plus

Planificateur de ressources distribué de stockage VMware

VMware Storage Distributed Resource Scheduler (SDRS) est une fonctionnalité vSphere qui place les machines virtuelles sur un stockage en fonction de la latence d'E/S actuelle et de l'utilisation de l'espace. Il déplace ensuite la machine virtuelle ou les VMDK sans interruption entre les datastores d'un cluster de datastores (également appelé pod), en sélectionnant le meilleur datastore pour placer la machine virtuelle ou les VMDK dans le cluster de datastore. Un cluster de data stores est un ensemble de datastores similaires agrégés dans une unité de consommation unique du point de vue de l'administrateur vSphere.

Lorsque vous utilisez DES DTS avec les outils ONTAP pour VMware vSphere, vous devez d'abord créer un datastore avec le plug-in, utiliser vCenter pour créer le cluster de datastores, puis y ajouter le datastore. Une fois le cluster datastore créé, des datastores supplémentaires peuvent être ajoutés au cluster datastore directement à partir de l'assistant de provisionnement sur la page Détails.

Les autres meilleures pratiques ONTAP en matière DE SDRS sont les suivantes :

- Tous les datastores du cluster doivent utiliser le même type de stockage (SAS, SATA ou SSD, par exemple), être tous des datastores VMFS ou NFS et disposer des mêmes paramètres de réplication et de protection.
- Envisagez d'utiliser DES DTS en mode par défaut (manuel). Cette approche vous permet d'examiner les recommandations et de décider s'il faut les appliquer ou non. Notez les effets suivants des migrations VMDK :
  - Lorsque DES DTS déplacent des VMDK entre les datastores, les économies d'espace éventuelles obtenues grâce au clonage ou à la déduplication ONTAP sont perdues. Vous pouvez réexécuter la déduplication pour récupérer ces économies.
  - Une fois que les DTS ont déplacé les VMDK, NetApp recommande de recréer les snapshots au niveau du datastore source car l'espace est autrement verrouillé par la machine virtuelle déplacée.
  - Le déplacement des VMDK entre les datastores du même agrégat n'a que peu d'avantages et LES DTS n'ont pas de visibilité sur d'autres charges de travail qui pourraient partager l'agrégat.

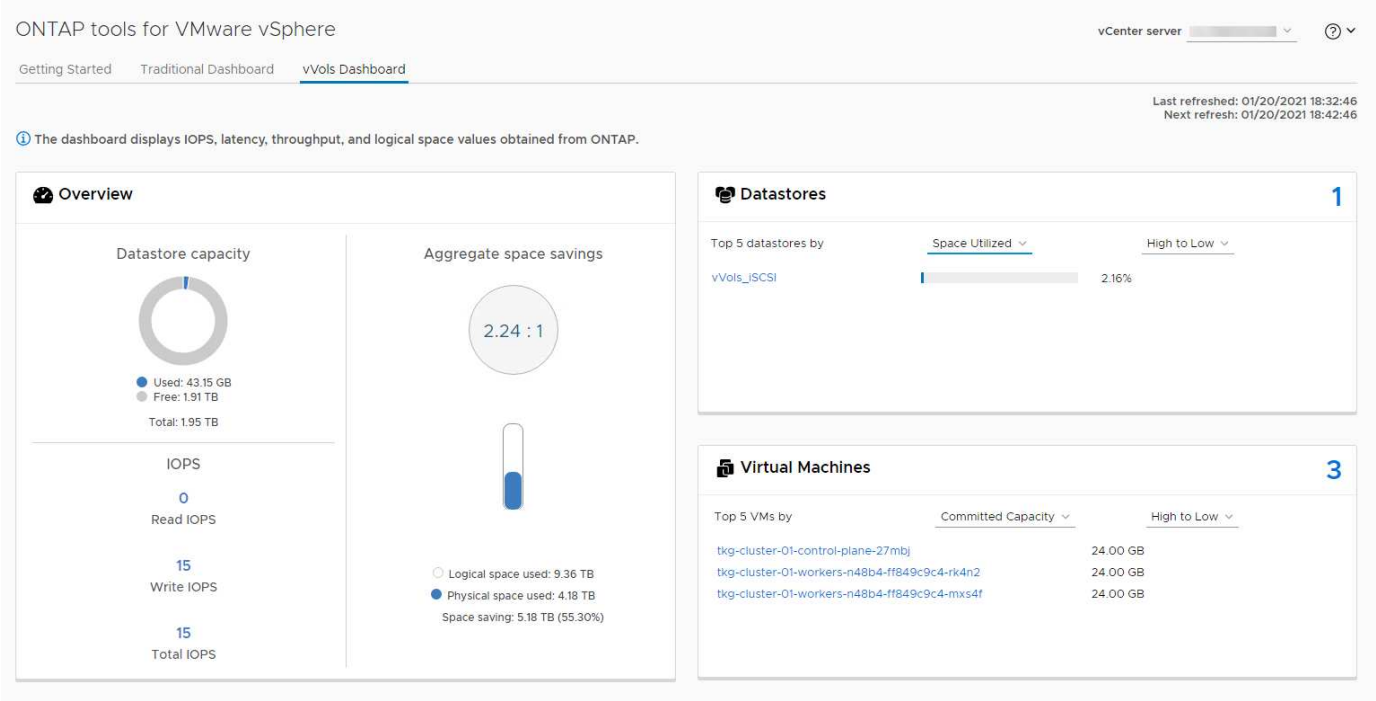
Gestion basée sur des règles de stockage et vVols

VMware vSphere APIs for Storage Awareness (VASA) permet aux administrateurs du stockage de configurer facilement des datastores avec des fonctionnalités bien définies et à l'administrateur des VM de les utiliser lorsque cela est nécessaire pour provisionner des VM sans avoir à interagir les uns avec les autres. Il est intéressant d'étudier cette approche pour savoir comment rationaliser vos opérations de stockage de virtualisation et éviter un travail insignifiant.

Avant VASA, les administrateurs des VM pouvaient définir des règles de stockage des VM, mais ils devaient travailler avec l'administrateur du stockage pour identifier les datastores appropriés, souvent à l'aide de la documentation ou des conventions de nommage. Grâce à VASA, l'administrateur du stockage peut définir un éventail de fonctionnalités de stockage, notamment la performance, le Tiering, le chiffrement et la réplication. Un ensemble de capacités pour un volume ou un ensemble de volumes est appelé « profil de capacité de stockage » (SCP).

Le SCP prend en charge la QoS minimale et/ou maximale pour les vVols de données d'une machine virtuelle. La QoS minimale est prise en charge uniquement sur les systèmes AFF. Les outils ONTAP pour VMware vSphere comprennent un tableau de bord affichant des performances granulaires de machine virtuelle et une capacité logique pour vVvols sur les systèmes ONTAP.

La figure suivante représente le tableau de bord des outils ONTAP pour VMware vSphere 9.8 vvols.



Une fois le profil de capacité de stockage défini, il peut être utilisé pour provisionner les machines virtuelles à l'aide de la règle de stockage qui identifie ses exigences. Le mappage entre la stratégie de stockage de la machine virtuelle et le profil de capacité de stockage du datastore permet à vCenter d'afficher la liste des datastores compatibles à sélectionner. Cette approche est appelée gestion basée sur des règles de stockage.

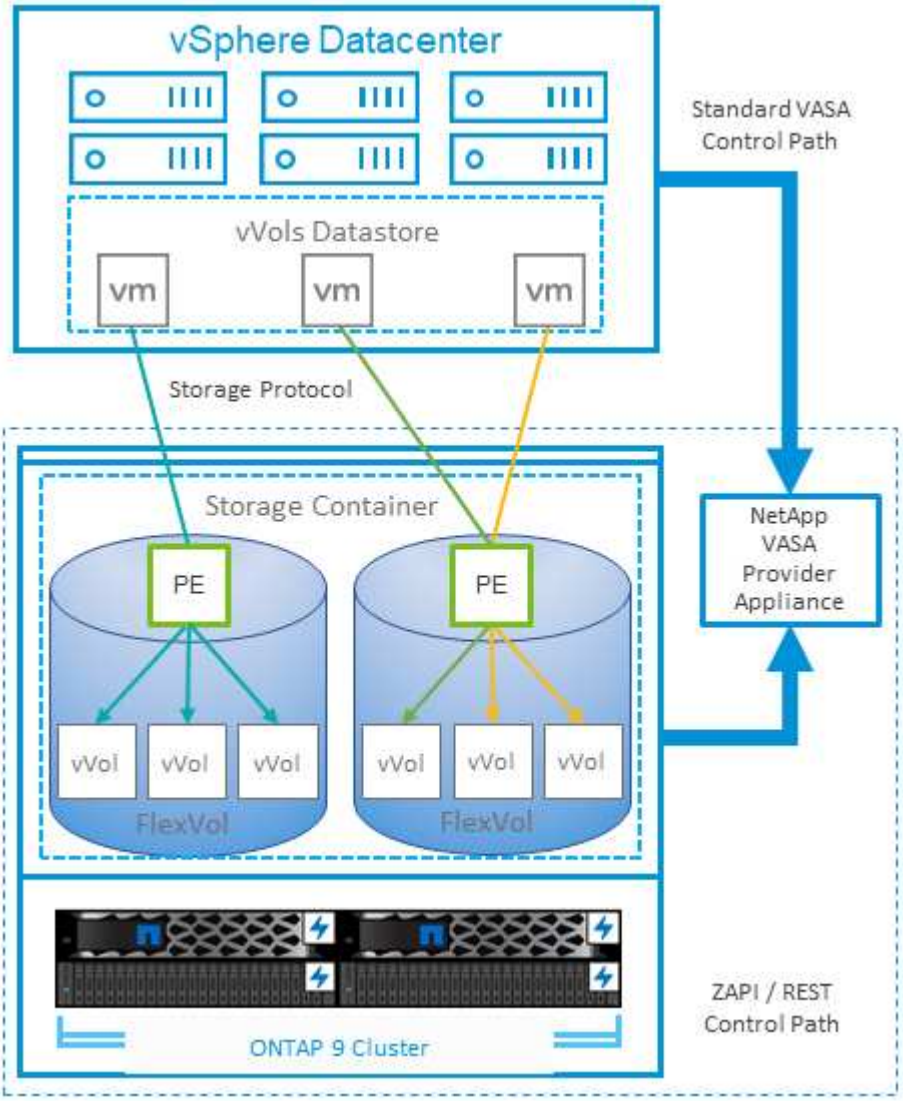
Vasa fournit la technologie permettant d'interroger le stockage et de renvoyer un ensemble de fonctionnalités de stockage vers vCenter. Les fournisseurs de VASA fournissent la traduction entre les API et les constructions du système de stockage et les API VMware que vCenter comprend. Le fournisseur VASA de NetApp pour ONTAP est proposé dans le cadre des outils ONTAP pour la machine virtuelle de l'appliance VMware vSphere. Le plug-in vCenter fournit l'interface de provisionnement et de gestion des datastores vVol, ainsi que la possibilité de définir des profils SCP (Storage Capability Profiles).

ONTAP prend en charge les datastores VMFS et NFS vvol. L'utilisation de vvols avec des datastores SAN apporte certains des avantages de NFS tels que la granularité au niveau des VM. Voici quelques meilleures pratiques à prendre en compte, et vous trouverez des informations supplémentaires dans le "TR-4400":

- Un datastore vvol peut être constitué de plusieurs volumes FlexVol sur plusieurs nœuds de cluster. L'approche la plus simple est un datastore unique, même si les volumes ont des capacités différentes. Grâce à la gestion du stockage basée sur des règles, un volume compatible est utilisé pour la machine virtuelle. Cependant, ces volumes doivent tous faire partie d'un seul SVM ONTAP et être accessibles via un seul protocole. Une LIF par nœud suffit pour chaque protocole. Évitez d'utiliser plusieurs versions de ONTAP dans un datastore vvol unique car les capacités de stockage peuvent varier d'une version à l'autre.
- Utilisez les outils ONTAP pour le plug-in VMware vSphere pour créer et gérer des datastores vvol. En plus de gérer le datastore et son profil, il crée automatiquement un terminal de protocole permettant d'accéder aux vvols si nécessaire. Si les LUN sont utilisées, notez que les terminaux PE sont mappés à l'aide des ID de LUN 300 et supérieurs. Vérifiez que le paramètre système avancé de l'hôte ESXi est défini `Disk.MaxLUN` Autorise un ID de LUN supérieur à 300 (la valeur par défaut est 1,024). Pour ce faire, sélectionnez l'hôte ESXi dans vCenter, puis l'onglet configurer et Rechercher `Disk.MaxLUN` Dans la liste des paramètres système avancés.
- N'installez pas ni ne migrez de VASA Provider, vCenter Server (appliance ou base Windows), ou les outils ONTAP pour VMware vSphere lui-même vers un datastore vvols, car ils sont ensuite interdépendants et limitent votre capacité à les gérer en cas de panne de courant ou d'autre perturbation du data Center.

- Sauvegarder régulièrement la machine virtuelle de VASA Provider. Créez au moins des copies Snapshot toutes les heures du datastore classique contenant VASA Provider. Pour en savoir plus sur la protection et la restauration de VASA Provider, consultez cette section "Article de la base de connaissances".

La figure suivante montre les composants de vvols.



Migration et sauvegarde dans le cloud

ONTAP permet également la prise en charge étendue du cloud hybride en fusionnant les systèmes de votre cloud privé sur site avec des capacités de cloud public. Voici quelques solutions clouds NetApp qui peuvent être utilisées en association avec vSphere :

- **Offres de première partie.** Amazon FSX pour NetApp ONTAP, Google Cloud NetApp volumes et Azure NetApp Files offrent des services de stockage gérés multiprotocole haute performance dans les principaux environnements de cloud public. Elles peuvent être utilisées directement par VMware Cloud on AWS(VMC on AWS), Azure VMware solution(AVS) et Google Cloud VMware Engine(GCVE) en tant que datastores ou stockage pour les systèmes d'exploitation invités (GOS) et les instances de calcul.
- **Cloud Volumes ONTAP.** le logiciel de gestion des données NetApp Cloud Volumes ONTAP permet de contrôler et de protéger les données et d'optimiser l'efficacité du stockage, tout en bénéficiant de la flexibilité du cloud de votre choix. Cloud Volumes ONTAP est un logiciel de gestion des données cloud

basé sur le stockage ONTAP. Utilisez-les conjointement avec Cloud Manager pour déployer et gérer des instances Cloud Volumes ONTAP avec vos systèmes ONTAP sur site. Profitez des fonctionnalités NAS avancées et SAN iSCSI combinées à la gestion unifiée des données, notamment les copies Snapshot et la réplication SnapMirror.

- **Services Cloud.** Utilisez la sauvegarde et la restauration BlueXP ou le cloud SnapMirror pour protéger les données des systèmes sur site via le stockage dans le cloud public. Cloud Sync vous aide à migrer et à synchroniser vos données entre le NAS, les magasins d'objets et le stockage Cloud Volumes Service. La reprise d'activité BlueXP est économique et efficace et permet d'exploiter les technologies NetApp comme base d'une solution de reprise d'activité robuste et efficace pour la reprise d'activité dans le cloud, la reprise d'activité sur site et sur site vers un environnement sur site.
- **FabricPool.** FabricPool offre un Tiering simple et rapide pour les données ONTAP. Les blocs inactifs peuvent être migrés vers un magasin d'objets dans des clouds publics ou un magasin d'objets StorageGRID privé. Ils sont automatiquement rappelés lorsque vous accédez de nouveau aux données ONTAP. Vous pouvez également utiliser le Tier objet comme troisième niveau de protection pour les données déjà gérées par SnapVault. Cette approche peut vous permettre de "Stockez davantage de snapshots de vos machines virtuelles" Sur les systèmes de stockage ONTAP primaires et/ou secondaires.
- **ONTAP Select.** utilisez le stockage Software-defined NetApp pour étendre votre cloud privé sur Internet aux sites et bureaux distants, où vous pouvez utiliser ONTAP Select pour prendre en charge les services de blocs et de fichiers ainsi que les mêmes fonctionnalités de gestion de données vSphere que votre data Center d'entreprise.

Lors de la conception de vos applications basées sur une VM, pensez à la mobilité future du cloud. Par exemple, plutôt que de placer les fichiers d'application et de données en même temps que les fichiers de données, utilisez une exportation LUN ou NFS distincte. Cela vous permet de migrer la machine virtuelle et les données séparément vers des services cloud.

Pour en savoir plus sur la sécurité, consultez ces ressources.

- "Documentation Cloud Volumes ONTAP"
- "Documentation ONTAP Select"
- "Documentation sur la sauvegarde et la restauration BlueXP"
- "Documentation sur la reprise d'activité BlueXP"
- "Amazon FSX pour NetApp ONTAP"
- "VMware Cloud sur AWS"
- "Qu'est-ce que Azure NetApp Files ?
"
- "Solution Azure VMware"
- "Moteur VMware Google Cloud"
- "Qu'est-ce que Google Cloud NetApp volumes ?"

Chiffrement pour les données vSphere

Aujourd'hui, les exigences croissantes en matière de protection des données au repos sont liées au chiffrement. Bien que la priorité initiale ait été donnée aux informations financières et de santé, il est de plus en plus intéressant de protéger toutes les informations, qu'elles soient stockées dans des fichiers, des bases de données ou tout autre type de données.

Les systèmes qui exécutent ONTAP simplifient la protection de toutes les données au moyen du chiffrement des données au repos. NetApp Storage Encryption (NSE) utilise les disques à autochiffrement (SED) avec ONTAP pour protéger les données SAN et NAS. NetApp propose également NetApp Volume Encryption et NetApp Aggregate Encryption comme une approche logicielle simple pour le chiffrement des volumes sur tous les disques. Ce chiffrement logiciel ne nécessite pas de disques spéciaux ni de gestionnaires de clés externes. Il est disponible gratuitement pour les clients ONTAP. Vous pouvez procéder à la mise à niveau et commencer à l'utiliser sans perturber vos clients ou vos applications. Elles sont validées par la norme FIPS 140-2 de niveau 1, y compris le gestionnaire de clés intégré.

Il existe plusieurs approches de protection des données des applications virtualisées qui s'exécutent sur VMware vSphere. L'une d'elles consiste à protéger les données avec les logiciels internes à la machine virtuelle au niveau du système d'exploitation invité. Les nouveaux hyperviseurs, tels que vSphere 6.5, prennent désormais en charge le cryptage au niveau des machines virtuelles. Cependant, le chiffrement logiciel NetApp est simple et facile :

- **Aucun effet sur la CPU du serveur virtuel.** certains environnements de serveurs virtuels nécessitent chaque cycle CPU disponible pour leurs applications, mais les tests ont montré que jusqu'à 5x ressources CPU sont nécessaires avec le cryptage au niveau de l'hyperviseur. Même si le logiciel de chiffrement prend en charge l'ensemble d'instructions AES-ni d'Intel pour décharger la charge de travail de chiffrement (comme le fait le chiffrement du logiciel NetApp), cette approche peut ne pas être possible en raison de l'exigence de nouveaux processeurs non compatibles avec les anciens serveurs.
- **Gestionnaire de clés intégré inclus.** Le chiffrement logiciel NetApp inclut un gestionnaire de clés intégré, sans frais supplémentaires. Vous pouvez ainsi vous lancer facilement sans serveurs de gestion des clés haute disponibilité complexes à l'achat et à l'utilisation.
- **Aucun effet sur l'efficacité du stockage.** les techniques d'efficacité du stockage comme la déduplication et la compression sont largement utilisées aujourd'hui et sont essentielles pour exploiter les supports disque Flash de façon rentable. Toutefois, les données cryptées ne sont en général pas dédupliquées ou compressées. Le cryptage du stockage et du matériel NetApp fonctionne à un niveau inférieur et permet l'utilisation totale des fonctionnalités d'efficacité du stockage NetApp, contrairement aux autres approches.
- **Chiffrement granulaire simple des datastores.** avec NetApp Volume Encryption, chaque volume bénéficie de sa propre clé AES 256 bits. Si vous devez le modifier, utilisez une seule commande. Cette approche est idéale si vous disposez de plusieurs locataires ou si vous devez prouver votre chiffrement indépendant pour différents services ou applications. Ce chiffrement est géré au niveau du datastore, ce qui est bien plus simple que de gérer des machines virtuelles individuelles.

La prise en main du chiffrement logiciel est très simple. Une fois la licence installée, configurez simplement le gestionnaire de clés intégré en spécifiant une phrase de passe, puis créez un nouveau volume ou déplacez le volume côté stockage pour activer le chiffrement. NetApp travaille à ajouter une prise en charge plus intégrée des fonctionnalités de cryptage dans les prochaines versions de ses outils VMware.

Pour en savoir plus sur la sécurité, consultez ces ressources.

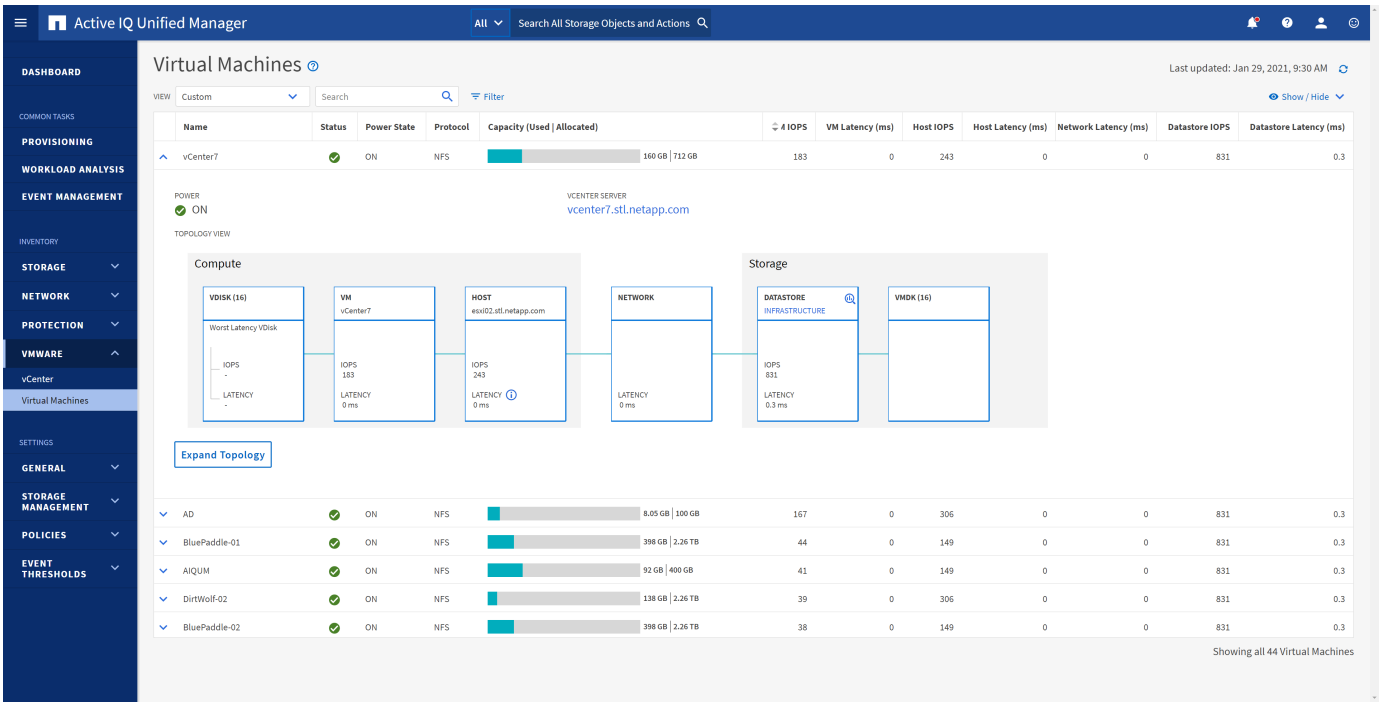
- "Rapports techniques sur la sécurité"
- "Guides de renforcement de la sécurité"
- "La documentation produit relative à la sécurité et au chiffrement des données ONTAP"

Active IQ Unified Manager

Active IQ Unified Manager permet d'avoir une grande visibilité sur les machines virtuelles de votre infrastructure virtuelle et assure la surveillance et le dépannage des problèmes de stockage et de performances dans votre environnement virtuel.

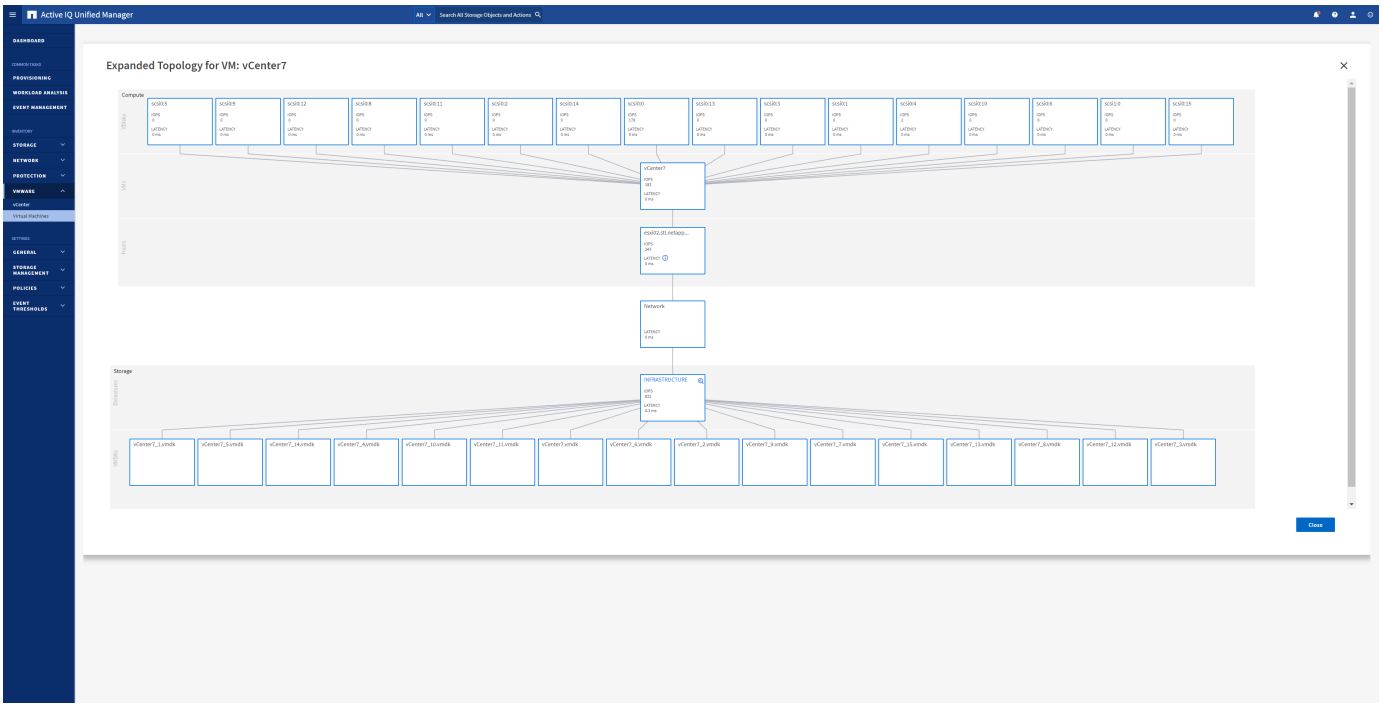
Un déploiement d'infrastructure virtuelle standard sur ONTAP comporte divers composants répartis sur les couches de calcul, de réseau et de stockage. Tout ralentissement des performances dans une application VM peut survenir en raison de la combinaison de latences rencontrées par les différents composants au niveau des couches respectives.

La capture d'écran suivante présente la vue des machines virtuelles Active IQ Unified Manager.



Unified Manager présente le sous-système sous-jacent d'un environnement virtuel dans une vue topologique afin de déterminer si un problème de latence a eu lieu dans le nœud de calcul, le réseau ou le stockage. La vue indique également l'objet spécifique qui provoque le décalage des performances lors de la réalisation des étapes correctives et de la résolution du problème sous-jacent.

La capture d'écran suivante montre la topologie étendue AIQUM.



Gestion basée sur des règles de stockage et vVols

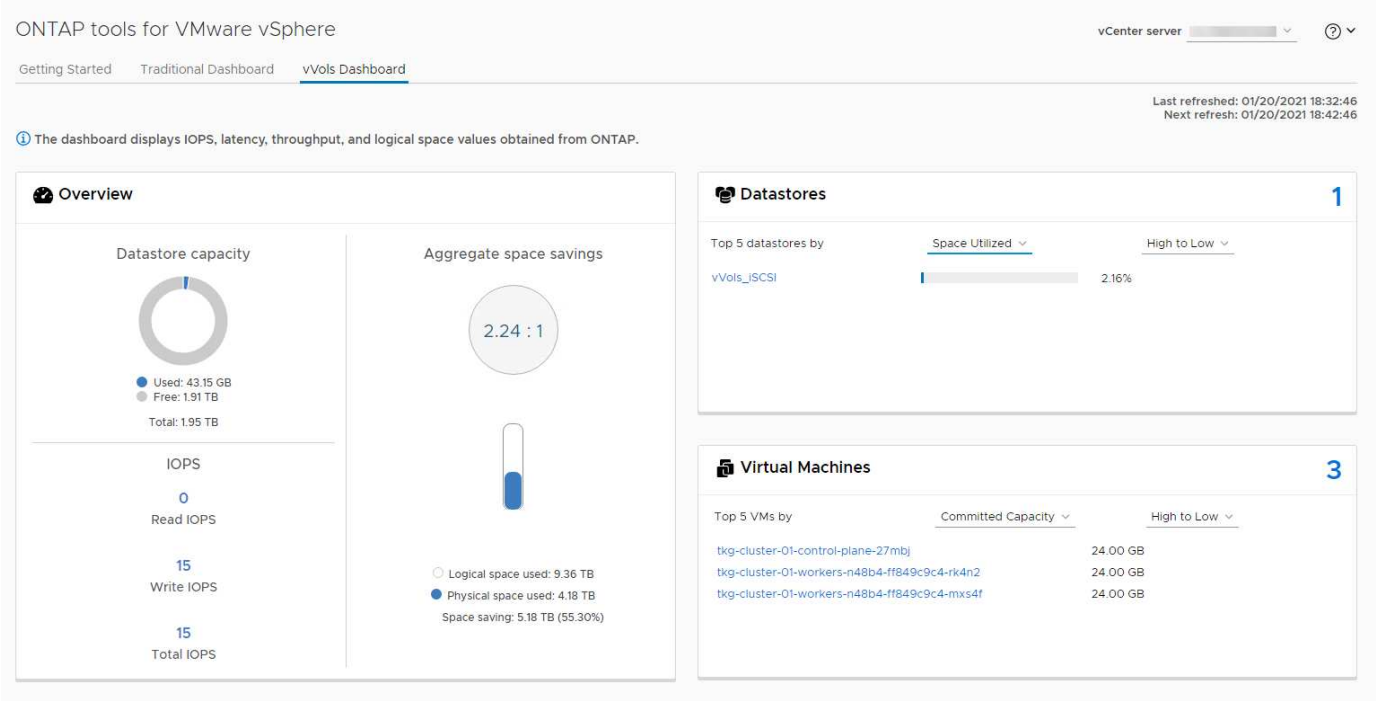
VMware vSphere APIs for Storage Awareness (VASA) permet aux administrateurs du stockage de configurer facilement des datastores avec des fonctionnalités bien définies et à l'administrateur des VM de les utiliser lorsque cela est nécessaire pour provisionner des VM sans avoir à interagir les uns avec les autres.

Il est intéressant d'étudier cette approche pour savoir comment rationaliser vos opérations de stockage de virtualisation et éviter un travail insignifiant.

Avant VASA, les administrateurs des VM pouvaient définir des règles de stockage des VM, mais ils devaient travailler avec l'administrateur du stockage pour identifier les datastores appropriés, souvent à l'aide de la documentation ou des conventions de nommage. Grâce à VASA, l'administrateur du stockage peut définir un éventail de fonctionnalités de stockage, notamment la performance, le Tiering, le chiffrement et la réplication. Un ensemble de capacités pour un volume ou un ensemble de volumes est appelé « profil de capacité de stockage » (SCP).

Le SCP prend en charge la QoS minimale et/ou maximale pour les vVols de données d'une machine virtuelle. La QoS minimale est prise en charge uniquement sur les systèmes AFF. Les outils ONTAP pour VMware vSphere comprennent un tableau de bord affichant des performances granulaires de machine virtuelle et une capacité logique pour vVvols sur les systèmes ONTAP.

La figure suivante représente le tableau de bord des outils ONTAP pour VMware vSphere 9.8 vvols.



Une fois le profil de capacité de stockage défini, il peut être utilisé pour provisionner les machines virtuelles à l'aide de la règle de stockage qui identifie ses exigences. Le mappage entre la stratégie de stockage de la machine virtuelle et le profil de capacité de stockage du datastore permet à vCenter d'afficher la liste des datastores compatibles à sélectionner. Cette approche est appelée gestion basée sur des règles de stockage.

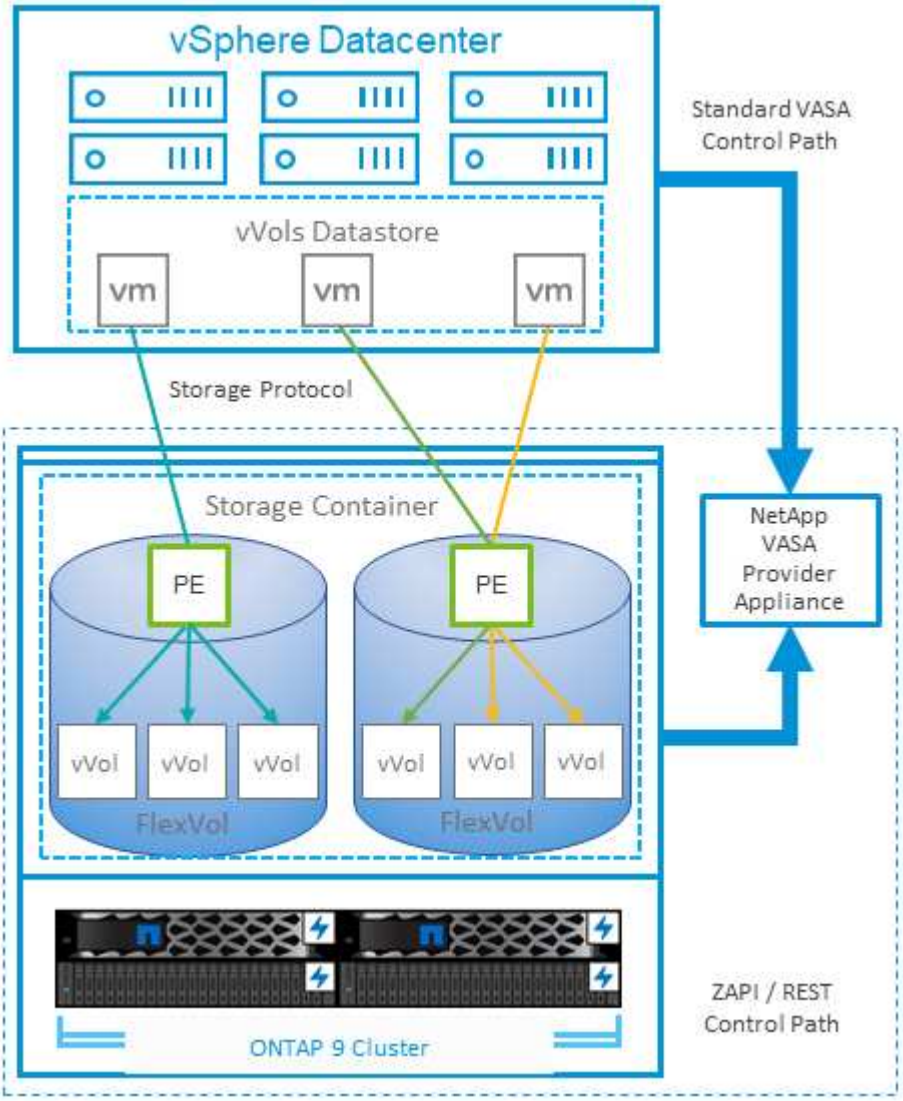
Vasa fournit la technologie permettant d'interroger le stockage et de renvoyer un ensemble de fonctionnalités de stockage vers vCenter. Les fournisseurs de VASA fournissent la traduction entre les API et les constructions du système de stockage et les API VMware que vCenter comprend. Le fournisseur VASA de NetApp pour ONTAP est proposé dans le cadre des outils ONTAP pour la machine virtuelle de l'appliance VMware vSphere. Le plug-in vCenter fournit l'interface de provisionnement et de gestion des datastores vVol, ainsi que la possibilité de définir des profils SCP (Storage Capability Profiles).

ONTAP prend en charge les datastores VMFS et NFS vvol. L'utilisation de vvols avec des datastores SAN apporte certains des avantages de NFS tels que la granularité au niveau des VM. Voici quelques meilleures pratiques à prendre en compte, et vous trouverez des informations supplémentaires dans le "TR-4400":

- Un datastore vvol peut être constitué de plusieurs volumes FlexVol sur plusieurs nœuds de cluster. L'approche la plus simple est un datastore unique, même si les volumes ont des capacités différentes. Grâce à la gestion du stockage basée sur des règles, un volume compatible est utilisé pour la machine virtuelle. Cependant, ces volumes doivent tous faire partie d'un seul SVM ONTAP et être accessibles via un seul protocole. Une LIF par nœud suffit pour chaque protocole. Évitez d'utiliser plusieurs versions de ONTAP dans un datastore vvol unique car les capacités de stockage peuvent varier d'une version à l'autre.
- Utilisez les outils ONTAP pour le plug-in VMware vSphere pour créer et gérer des datastores vvol. En plus de gérer le datastore et son profil, il crée automatiquement un terminal de protocole permettant d'accéder aux vvols si nécessaire. Si les LUN sont utilisées, notez que les terminaux PE sont mappés à l'aide des ID de LUN 300 et supérieurs. Vérifiez que le paramètre système avancé de l'hôte ESXi est défini `Disk.MaxLUN` Autorise un ID de LUN supérieur à 300 (la valeur par défaut est 1,024). Pour ce faire, sélectionnez l'hôte ESXi dans vCenter, puis l'onglet configurer et Rechercher `Disk.MaxLUN` Dans la liste des paramètres système avancés.
- N'installez pas ni ne migrez de VASA Provider, vCenter Server (appliance ou base Windows), ou les outils ONTAP pour VMware vSphere lui-même vers un datastore vvols, car ils sont ensuite interdépendants et limitent votre capacité à les gérer en cas de panne de courant ou d'autre perturbation du data Center.

- Sauvegarder régulièrement la machine virtuelle de VASA Provider. Créez au moins des copies Snapshot toutes les heures du datastore classique contenant VASA Provider. Pour en savoir plus sur la protection et la restauration de VASA Provider, consultez cette section "Article de la base de connaissances".

La figure suivante montre les composants de vvols.



Planificateur de ressources distribué de stockage VMware

VMware Storage Distributed Resource Scheduler (SDRS) est une fonction vSphere qui place automatiquement les machines virtuelles dans un cluster de datastores en fonction de la latence d'E/S actuelle et de l'utilisation de l'espace.

Il déplace ensuite la machine virtuelle ou les VMDK sans interruption entre les datastores d'un cluster de datastores (également appelé pod), en sélectionnant le meilleur datastore pour placer la machine virtuelle ou les VMDK dans le cluster de datastore. Un cluster de data stores est un ensemble de datastores similaires agrégés dans une unité de consommation unique du point de vue de l'administrateur vSphere.

Lorsque vous utilisez DES DTS avec les outils ONTAP pour VMware vSphere, vous devez d'abord créer un datastore avec le plug-in, utiliser vCenter pour créer le cluster de datastores, puis y ajouter le datastore. Une fois le cluster datastore créé, des datastores supplémentaires peuvent être ajoutés au cluster datastore

directement à partir de l'assistant de provisionnement sur la page Détails.

Les autres meilleures pratiques ONTAP en matière DE SDRS sont les suivantes :

- N'utilisez pas DE DTS à moins d'avoir une exigence spécifique pour le faire.
  - LES DTS ne sont pas nécessaires lors de l'utilisation de ONTAP. LES SDRS n'ont pas connaissance des fonctionnalités d'efficacité du stockage ONTAP, telles que la déduplication et la compression, et peuvent donc prendre des décisions qui ne sont pas optimales pour votre environnement.
  - LES DTS n'ont pas connaissance des règles de QoS de ONTAP et peuvent donc prendre des décisions qui ne sont pas optimales pour la performance.
  - LES DTS ne connaissent pas les copies snapshot ONTAP et peuvent donc prendre des décisions qui entraînent une croissance exponentielle des snapshots. Par exemple, le déplacement d'une machine virtuelle vers un autre datastore crée de nouveaux fichiers dans le nouveau datastore, ce qui entraîne l'augmentation du snapshot. Cela est particulièrement vrai pour les machines virtuelles équipées de disques de grande taille ou de nombreux snapshots. Ensuite, si la machine virtuelle doit être replacée dans le datastore d'origine, le snapshot du datastore d'origine augmentera encore plus.

Si vous utilisez DES DTS, tenez compte des meilleures pratiques suivantes :

- Tous les datastores du cluster doivent utiliser le même type de stockage (SAS, SATA ou SSD, par exemple), être tous des datastores VMFS ou NFS et disposer des mêmes paramètres de réplication et de protection.
- Envisagez d'utiliser DES DTS en mode par défaut (manuel). Cette approche vous permet d'examiner les recommandations et de décider s'il faut les appliquer ou non. Notez les effets suivants des migrations VMDK :
  - Lorsque LES DTS déplacent des VMDK entre les datastores, les économies d'espace éventuelles liées au clonage ou à la déduplication ONTAP peuvent être réduites selon la qualité de déduplication ou de compression sur la destination.
  - Une fois que les DTS ont déplacé les VMDK, NetApp recommande de recréer les snapshots au niveau du datastore source car l'espace est autrement verrouillé par la machine virtuelle déplacée.
  - Le déplacement des VMDK entre les datastores du même agrégat n'a que peu d'avantages et LES DTS n'ont pas de visibilité sur d'autres charges de travail qui pourraient partager l'agrégat.

Pour plus d'informations sur les DTS, consultez la documentation VMware à l'adresse "FAQ sur Storage DRS".

Hôte ESXi recommandé et autres paramètres ONTAP recommandés

NetApp a développé un ensemble de paramètres hôtes ESXi optimaux pour les protocoles NFS et les protocoles en mode bloc. Des conseils spécifiques sont également fournis concernant les paramètres de chemins d'accès multiples et de délai d'expiration des HBA pour un comportement correct avec ONTAP basé sur les tests internes NetApp et VMware.

Ces valeurs sont facilement définies à l'aide des outils ONTAP pour VMware vSphere : dans la page de présentation des outils ONTAP, faites défiler vers le bas et cliquez sur appliquer les paramètres recommandés dans le portlet conformité des hôtes ESXi.

Voici les paramètres d'hôte recommandés pour toutes les versions de ONTAP actuellement prises en charge.

Paramètres hôte	Valeur recommandée par NetApp	Redémarrer requis
Configuration avancée ESXi		
VMFS3.HardwareAccélérationde la localisation	Conserver la valeur par défaut (1)	Non
VMFS3.EnableBlockDelete	Conserver la valeur par défaut (0), mais peut être modifiée si nécessaire. Pour plus d'informations, voir "Récupération d'espace pour les machines virtuelles VMFS5"	Non
VMFS3.EnableVMFS6Unmap	Conserver la valeur par défaut (1) pour plus d'informations, voir "API VMware vSphere : intégration des baies (VAAI)"	Non
Paramètres NFS		
NewSyncInterval	Si vous n'utilisez pas vSphere CSI pour Kubernetes, définissez-le comme indiqué "VMware KB 386364"	Non
Net.TcpipeHeapSize	VSphere 6.0 ou version ultérieure, défini sur 32. Toutes les autres configurations NFS, définies sur 30	Oui.
Net.TcpipeHeapMax	Défini sur 512 Mo pour la plupart des versions vSphere 6.X. Réglez sur la valeur par défaut (1024 Mo) pour 6.5U3, 6.7U3 et 7.0 ou ultérieure.	Oui.
NFS.MaxVolumes	VSphere 6.0 ou version ultérieure, défini sur 256 Toutes les autres configurations NFS définies sur 64.	Non
NFS41.Maxvolumes	VSphere 6.0 ou version ultérieure, défini sur 256.	Non
NFS.MaxQueueDepth[1]	VSphere 6.0 ou version ultérieure, défini sur 128	Oui.
NFS.HeartbeatMaxFailures	Définissez sur 10 pour l'ensemble des configurations NFS	Non
NFS.HeartbeatFrequency	Définissez la valeur 12 pour toutes les configurations NFS	Non
NFS.HeartbeatTimeout	Définissez sur 5 pour l'ensemble des configurations NFS.	Non
Sunrpc.MaxConnPerIP	vSphere 7.0 à 8.0, défini sur 128. Ce paramètre est ignoré dans les versions ESXi ultérieures à 8.0.	Non

Paramètres hôte	Valeur recommandée par NetApp	Redémarrer requis
Paramètres FC/FCoE		
Stratégie de sélection de chemin	Définissez-le sur RR (Round Robin) lorsque des chemins FC avec ALUA sont utilisés. Défini sur FIXE pour toutes les autres configurations. La définition de cette valeur sur RR permet d'équilibrer la charge sur l'ensemble des chemins actifs/optimisés. La valeur FIXÉE est pour les anciennes configurations non ALUA et contribue à empêcher les E/S proxy En d'autres termes, il contribue à empêcher les E/S de se diriger vers l'autre nœud d'une paire haute disponibilité dans un environnement doté de Data ONTAP 7-mode	Non
Disk.QFullSampleSize	Définissez sur 32 pour toutes les configurations. La définition de cette valeur permet d'éviter les erreurs d'E/S.	Non
Disk.QFullThreshold	Réglez à 8 pour toutes les configurations. La définition de cette valeur permet d'éviter les erreurs d'E/S.	Non
Délais d'expiration de la carte HBA FC Emulex	Utilisez la valeur par défaut.	Non
Délais de connexion HBA FC QLogic	Utilisez la valeur par défaut.	Non
Paramètres iSCSI		
Stratégie de sélection de chemin	Définissez à RR (Round Robin) pour tous les chemins iSCSI. La définition de cette valeur sur RR permet d'équilibrer la charge sur l'ensemble des chemins actifs/optimisés.	Non
Disk.QFullSampleSize	Définissez sur 32 pour toutes les configurations. La définition de cette valeur permet d'éviter les erreurs d'E/S.	Non
Disk.QFullThreshold	Réglez à 8 pour toutes les configurations. La définition de cette valeur permet d'éviter les erreurs d'E/S.	Non

> Option de configuration avancée NFS MaxQueueDepth peut ne pas fonctionner comme prévu lors de l'utilisation de VMware vSphere ESXi 7.0.1 et de VMware vSphere ESXi 7.0.2. Référence "VMware KB 86331" pour plus d'informations.

Lors de la création de volumes et de LUN ONTAP FlexVol, les outils ONTAP permettent également de spécifier certains paramètres par défaut :

Outil ONTAP	Paramètre par défaut
Réserve Snapshot (-percent-snapshot-space)	0
Réserve fractionnaire (-réserve fractionnaire)	0
Mise à jour de l'heure d'accès (-atime-update)	Faux
Lecture minimum (-min-lecture anticipée)	Faux
Snapshots planifiés	Aucune
Efficacité du stockage	Activé
Garantie de volume	Aucune (provisionnement fin)
Taille automatique du volume	augmenter_réduire
Réservation d'espace par LUN	Désactivé
Allocation d'espace de la LUN	Activé

Paramètres de chemins d'accès multiples pour les performances

Bien qu'il ne soit pas actuellement configuré par les outils ONTAP disponibles, NetApp suggère les options de configuration suivantes :

- Lorsque vous utilisez des systèmes non ASA dans des environnements hautes performances ou lorsque vous testez les performances avec une seule banque de données LUN, envisagez de modifier le paramètre d'équilibrage de charge de la stratégie de sélection de chemin (PSP) à tour de rôle (VMW_PSP_RR) du paramètre IOPS par défaut de 1 000 à une valeur de 1. Voir"VMware KB 2069356" pour plus d'infos.
- Dans vSphere 6.7 Update 1, VMware a introduit un nouveau mécanisme d'équilibrage de charge de latence pour le Round Robin PSP. L'option de latence est désormais également disponible lors de l'utilisation du HPP (High Performance Plugin) avec les espaces de noms NVMe et avec vSphere 8.0u2 et versions ultérieures, les LUN connectés iSCSI et FCP. La nouvelle option prend en compte la bande passante d'E/S et la latence du chemin lors de la sélection du chemin optimal pour les E/S. NetApp recommande d'utiliser l'option de latence dans les environnements avec une connectivité de chemin non équivalente, comme dans les cas avec plus de sauts réseau sur un chemin que sur un autre, ou lors de l'utilisation d'un système NetApp ASA . Voir "Modifier les paramètres par défaut pour le tour de latence" pour plus d'informations.

Documentation complémentaire

Pour FCP et iSCSI avec vSphere 7, des informations supplémentaires sont disponibles à l'adresse "Utilisez VMware vSphere 7.x avec ONTAP" pour FCP et iSCSI avec vSphere 8. Vous trouverez plus de détails à l'adresse "Utilisez VMware vSphere 8.x avec ONTAP" concernant NVMe-of avec vSphere 7. Des informations plus détaillées sont disponibles à l'adresse "Pour plus de détails sur NVMe-of, consultez la page Configuration d'hôte NVMe-of pour ESXi 7.x avec ONTAP" concernant NVMe-of avec vSphere 8. Des informations plus détaillées sont disponibles à l'adresse "Pour plus de détails sur NVMe-of, consultez la page Configuration

Volumes virtuels (vVols) avec les outils ONTAP 10

Présentation

ONTAP est une solution de stockage leader pour les environnements VMware vSphere depuis plus de vingt ans et continue d’ajouter des fonctionnalités innovantes pour simplifier la gestion tout en réduisant les coûts.

Ce document présente les fonctionnalités de ONTAP pour les volumes virtuels VMware vSphere (vVols), notamment les dernières informations sur les produits et les cas d’utilisation, ainsi que les bonnes pratiques et d’autres informations permettant de rationaliser le déploiement et de réduire les erreurs.

> Cette documentation remplace les rapports techniques *TR-4400 : VMware vSphere Virtual volumes (vVols) par ONTAP*

Les meilleures pratiques complètent d’autres documents, tels que des guides et des listes de compatibilité. Ils sont développés en fonction de tests effectués en laboratoire et d’une vaste expérience sur le terrain par les ingénieurs et les clients NetApp. Ce ne sont peut-être pas les seules pratiques qui fonctionnent ou sont prises en charge, mais sont généralement les solutions les plus simples qui répondent aux besoins de la plupart des clients.

> Ce document a été mis à jour pour inclure les nouvelles fonctionnalités vVols de vSphere 8.0 mise à jour 3, la version 10.4 des outils ONTAP et les nouveaux systèmes NetApp ASA.

Présentation des volumes virtuels (vVols)

En 2012, NetApp a commencé à travailler avec VMware pour prendre en charge les API vSphere pour Storage Awareness (VASA) pour vSphere 5. Ce premier VASA Provider a autorisé la définition des fonctionnalités de stockage dans un profil qui pouvait être utilisé pour filtrer les datastores lors du provisionnement et pour vérifier par la suite la conformité avec la règle. Cette évolution a vu le jour, de nouvelles fonctionnalités permettant d’automatiser davantage le provisionnement, ainsi que l’ajout de volumes virtuels ou de vVols où des objets de stockage individuels sont utilisés pour les fichiers de machines virtuelles et les disques virtuels. Il peut s’agir de LUN, de fichiers et désormais d’espaces de noms vSphere 8 NVMe (utilisés avec les outils ONTAP 9.13P2). NetApp a travaillé en étroite collaboration avec VMware en tant que partenaire de référence pour les vVols publiés avec vSphere 6 en 2015, et à nouveau en tant que partenaire de conception pour les vVols utilisant NVMe over Fabrics dans vSphere 8. NetApp continue d’améliorer les vVols pour tirer parti des dernières fonctionnalités d’ONTAP.

Plusieurs composants doivent être pris en compte :

Vasa Provider

Il s’agit du composant logiciel qui gère la communication entre VMware vSphere et le système de stockage. Pour ONTAP, le fournisseur VASA s’exécute dans une appliance connue sous le nom d’outils ONTAP pour VMware vSphere (outils ONTAP pour, par exemple). Les outils ONTAP incluent également un plug-in vCenter, un adaptateur de réplication du stockage (SRA) pour VMware site Recovery Manager et un serveur d’API REST pour vous permettre de créer votre propre automatisation. Une fois les outils ONTAP configurés et enregistrés dans vCenter, il est désormais peu nécessaire d’interagir directement avec le système ONTAP, puisque la quasi-totalité de vos besoins en stockage peut être gérée directement depuis l’interface utilisateur vCenter ou via l’automatisation de l’API REST.

Terminal PE (Protocol Endpoint)

Le terminal de protocole est un proxy pour les E/S entre les hôtes ESXi et le datastore vVols. Le fournisseur ONTAP VASA les crée automatiquement, soit une LUN de terminal de protocole (4 Mo) par volume FlexVol du datastore vVols, soit un point de montage NFS par interface NFS (LIF) sur le nœud de stockage hébergeant un volume FlexVol dans le datastore. L'hôte ESXi monte ces terminaux de protocole directement plutôt que des LUN vVol individuelles et des fichiers de disque virtuel. Il n'est pas nécessaire de gérer les terminaux PE lorsqu'ils sont créés, montés, démontés et supprimés automatiquement par le fournisseur VASA, avec les groupes d'interfaces ou les règles d'exportation nécessaires.

Terminal virtuel de protocole (VPE)

Nouveauté de vSphere 8, lorsque NVMe over Fabrics (NVMe-of) avec vVols, le concept de terminal de protocole n'est plus pertinent dans ONTAP. Au lieu de cela, un PE virtuel est instancié automatiquement par l'hôte ESXi pour chaque groupe ANA dès que la première machine virtuelle est sous tension. ONTAP crée automatiquement des groupes ANA pour chaque volume FlexVol utilisé par le datastore.

Autre avantage de NVMe-of pour les vVols : aucune demande de liaison n'est requise du fournisseur VASA. À la place, l'hôte ESXi gère en interne la fonctionnalité de liaison vVol basée sur le VPE. Cela réduit les risques d'impact d'une tempête de liaison vVol sur le service.

Pour plus d'informations, voir "NVMe et les volumes virtuels" marche "vmware.com"

Datastore du volume virtuel

| le datastore de volume virtuel est une représentation de datastore logique d'un conteneur vVols créée et gérée par un fournisseur VASA. Le conteneur représente un pool de capacité de stockage provisionné à partir des systèmes de stockage gérés par le fournisseur VASA. Les outils ONTAP prennent en charge l'allocation de plusieurs volumes FlexVol (appelés « volumes de sauvegarde ») à un datastore vVols unique. Ces datastores vVols peuvent couvrir plusieurs nœuds dans un cluster ONTAP, combinant des systèmes Flash et hybrides ayant des fonctionnalités différentes. L'administrateur peut créer de nouveaux volumes FlexVol à l'aide de l'assistant de provisionnement ou de l'API REST, ou sélectionner des volumes FlexVol précréés pour la sauvegarde du stockage, le cas échéant.

Volumes virtuels (vVols)

VVols sont les fichiers et disques de machines virtuelles réellement stockés dans le datastore vVols. L'utilisation du terme vVol (singulier) fait référence à un fichier, une LUN ou un espace de nom spécifique unique. ONTAP crée des namespaces NVMe, des LUN ou des fichiers en fonction du protocole utilisé par le datastore. Il existe plusieurs types distincts de vVols : les plus courants sont Config (le seul avec VMFS, il contient des fichiers de métadonnées tels que le fichier VMX de la machine virtuelle), Data (disque virtuel ou VMDK) et Swap (créé lorsque la machine virtuelle est mise sous tension). Les vVols protégés par le chiffrement VMware VM seront de type autre. Le chiffrement des machines virtuelles VMware ne doit pas être confondu avec le chiffrement du volume ou de l'agrégat ONTAP.

Gestion stratégique

Avec VMware vSphere APIs for Storage Awareness (VASA), un administrateur de serveurs virtuels peut facilement utiliser les fonctionnalités de stockage nécessaires pour provisionner des serveurs virtuels sans avoir à interagir avec son équipe de stockage. Avant VASA, les administrateurs des serveurs virtuels pouvaient définir des règles de stockage des serveurs virtuels, mais ils devaient travailler avec leurs administrateurs du stockage pour identifier les datastores appropriés, souvent à l'aide de la documentation ou des conventions de nommage. Dans VASA, les administrateurs de vCenter disposant des autorisations appropriées peuvent définir une gamme de fonctionnalités de stockage que les utilisateurs de vCenter peuvent ensuite utiliser pour provisionner des VM. Le mappage entre la règle de stockage de la machine virtuelle et les fonctionnalités du

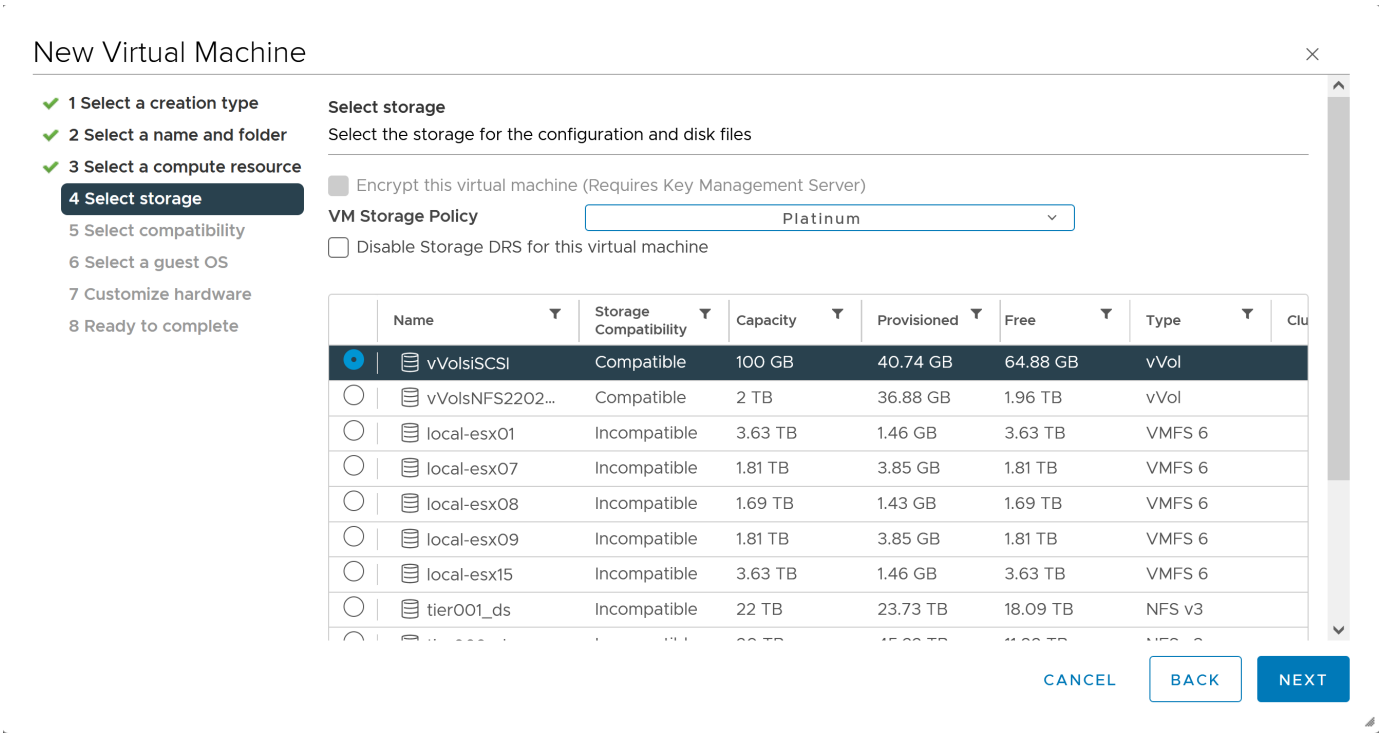
datastore permet à vCenter d'afficher une liste de datastores compatibles à sélectionner, ainsi que d'activer d'autres technologies telles que Aria (anciennement vRealize) Automation ou Tanzu Kubernetes Grid pour sélectionner automatiquement le stockage dans une règle attribuée. Cette approche est appelée gestion basée sur des règles de stockage. Si les règles VASA Provider et les politiques de stockage de VM peuvent également être utilisées avec les datastores traditionnels, notre attention se concentre sur les datastores vVols.

Règles de stockage de VM

Les règles de stockage de serveur virtuel sont créées dans vCenter sous stratégies et profils. Pour les vVols, créez un jeu de règles à l'aide de règles provenant du fournisseur de type de stockage NetApp vVols. Les outils ONTAP 10.X offrent désormais une approche plus simple que les outils ONTAP 9.X en vous permettant de spécifier directement les attributs de stockage dans la stratégie de stockage des machines virtuelles.

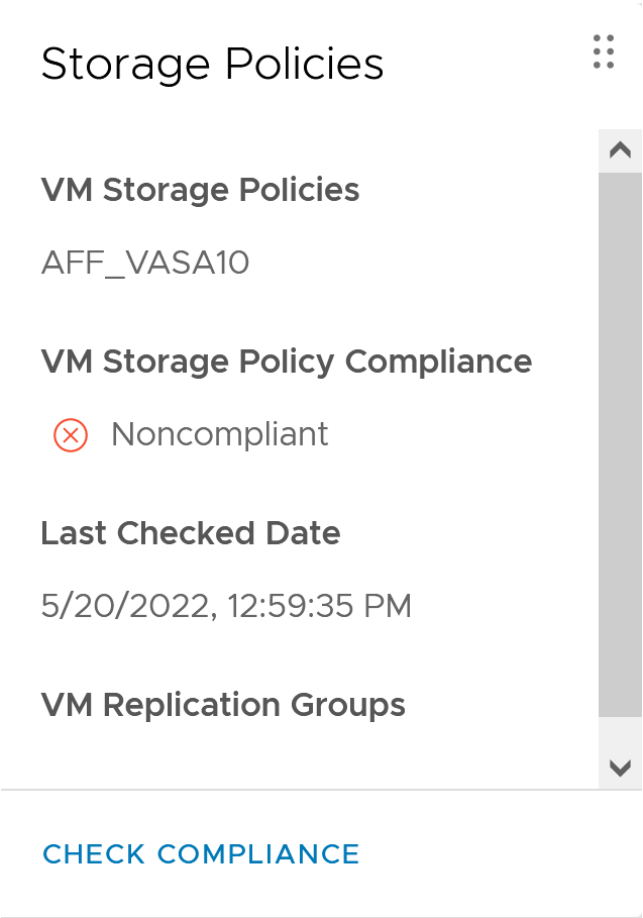
Comme mentionné ci-dessus, l'utilisation de règles peut aider à rationaliser la tâche de provisionnement d'une machine virtuelle ou d'un VMDK. Il vous suffit de sélectionner une règle appropriée, et le fournisseur VASA affiche les datastores vVols qui prennent en charge cette règle et place le vVol dans un FlexVol volume individuel conforme.

Déployer une machine virtuelle à l'aide de la stratégie de stockage



Une fois qu'une machine virtuelle est provisionnée, le fournisseur VASA continue à vérifier la conformité et alerte l'administrateur de la machine virtuelle en cas d'alarme dans vCenter lorsque le volume de sauvegarde n'est plus conforme à la règle.

Conformité à la règle de stockage VM



Prise en charge des vVols de NetApp

ONTAP prend en charge la spécification VASA depuis sa sortie initiale en 2012. Si d'autres systèmes de stockage NetApp peuvent prendre en charge VASA, ce document est axé sur les versions actuellement prises en charge de ONTAP 9.

ONTAP

Outre ONTAP 9 sur les systèmes AFF, ASA et FAS, NetApp prend en charge les workloads VMware sur ONTAP Select, Amazon FSX pour NetApp avec VMware Cloud sur AWS, Azure NetApp Files avec la solution Azure VMware, Cloud Volumes Service avec Google Cloud VMware Engine et le stockage privé NetApp dans Equinix, mais certaines fonctionnalités peuvent varier en fonction du fournisseur de services et de la connectivité réseau disponible. L'accès, depuis les invités vSphere, aux données stockées dans ces configurations ainsi qu'à Cloud Volumes ONTAP est également disponible.

Au moment de la publication, les environnements hyperscale sont limités aux datastores NFS v3 classiques. Par conséquent, les vVols ne sont disponibles que pour les systèmes ONTAP sur site ou les systèmes connectés au cloud qui offrent l'ensemble des fonctionnalités d'un système sur site, tels que ceux hébergés par les partenaires et fournisseurs de services NetApp à travers le monde.

Pour plus d'informations sur ONTAP, voir "Documentation des produits ONTAP"

Pour plus d'informations sur les meilleures pratiques ONTAP et VMware vSphere, voir "TR-4597"

Avantages de l'utilisation de vVols avec ONTAP

Lorsque VMware a introduit la prise en charge de vVols avec VASA 2.0 en 2015, ils l'ont décrite comme « une

structure d'intégration et de gestion fournissant un nouveau modèle opérationnel pour le stockage externe (SAN/NAS) ». Ce modèle opérationnel présente plusieurs avantages avec le stockage ONTAP.

Gestion stratégique

Comme décrit à la section 1.2, la gestion basée sur des règles permet de provisionner les machines virtuelles et de les gérer par la suite à l'aide de règles prédéfinies. Les opérations INFORMATIQUES peuvent ainsi être réalisées de plusieurs manières :

- **Augmentez la vitesse.** les outils ONTAP éliminent la nécessité pour l'administrateur vCenter d'ouvrir des tickets avec l'équipe chargée du stockage pour les activités de provisionnement du stockage. Cependant, les rôles RBAC des outils ONTAP dans vCenter et sur le système ONTAP permettent toujours l'accès à des équipes indépendantes (telles que les équipes chargées du stockage) ou à des activités indépendantes par la même équipe en limitant l'accès à des fonctions spécifiques si nécessaire.
- **Provisionnement plus intelligent.** les fonctionnalités du système de stockage peuvent être exposées via les API VASA, ce qui permet aux flux de travail de provisionnement de tirer parti de fonctionnalités avancées sans que l'administrateur des machines virtuelles ait besoin de comprendre comment gérer le système de stockage.
- **Provisionnement plus rapide.** différentes capacités de stockage peuvent être prises en charge dans un seul datastore et sélectionnées automatiquement comme approprié pour une machine virtuelle en fonction de la stratégie de la machine virtuelle.
- **Évitez les erreurs.** les stratégies de stockage et de machines virtuelles sont développées à l'avance et appliquées selon les besoins sans avoir à personnaliser le stockage à chaque fois qu'une machine virtuelle est provisionnée. Les alarmes de conformité sont déclenchées lorsque les fonctionnalités de stockage sont différentes des règles définies. Comme mentionné précédemment, les plateformes SCP rendent le provisionnement initial prévisible et reproductible, tandis que la base des règles de stockage des serveurs virtuels sur les plateformes SCP garantit un placement précis.
- **Meilleure gestion des capacités.** Les outils VASA et ONTAP permettent de consulter la capacité de stockage jusqu'au niveau des agrégats individuels si nécessaire et de fournir plusieurs couches d'alertes en cas de début d'exécution de la capacité.

Gestion granulaire des machines virtuelles dans le SAN moderne

Les systèmes DE stockage SAN utilisant Fibre Channel et iSCSI ont été les premiers à être pris en charge par VMware pour ESX, mais ils n'ont pas été en mesure de gérer les disques et les fichiers individuels des machines virtuelles à partir du système de stockage. Au lieu de cela, les LUN sont provisionnées et VMFS gère les fichiers individuels. Il est donc difficile pour le système de stockage de gérer directement les performances, le clonage et la protection du stockage des machines virtuelles individuelles. Les vVols apportent la granularité du stockage dont les clients utilisent déjà le stockage NFS, et les fonctionnalités SAN robustes et hautes performances de ONTAP.

Désormais, avec vSphere 8 et les outils ONTAP pour VMware vSphere 9.12 et versions ultérieures, les mêmes contrôles granulaires utilisés par les vVols pour les anciens protocoles SCSI sont désormais disponibles dans le SAN Fibre Channel moderne utilisant NVMe over Fabrics pour des performances encore plus élevées à grande échelle. Avec vSphere 8.0 mise à jour 1, il est désormais possible de déployer une solution NVMe de bout en bout complète à l'aide de vVols sans déplacement d'E/S dans la pile de stockage de l'hyperviseur.

Meilleures fonctionnalités de déchargement du stockage

Tandis que VAAI offre de nombreuses opérations qui sont déchargées vers le stockage, certaines lacunes sont traitées par le fournisseur VASA. SAN VAAI ne peut pas décharger les snapshots gérés par VMware vers le système de stockage. NFS VAAI peut décharger les snapshots gérés par les machines virtuelles, mais il existe des limites placées pour les machines virtuelles avec des snapshots natifs de stockage. Étant donné que les

vVols utilisent des LUN, des espaces de noms ou des fichiers individuels pour des disques de machines virtuelles, ONTAP peut rapidement et efficacement cloner les fichiers ou les LUN pour créer des snapshots granulaires de machines virtuelles qui ne nécessitent plus de fichiers delta. NFS VAAI ne prend pas non plus en charge les opérations de déchargement des clones pour les migrations Storage vMotion à chaud (basées sur). La machine virtuelle doit être mise hors tension pour permettre la décharge de la migration lors de l'utilisation de VAAI avec des datastores NFS classiques. Le fournisseur VASA des outils ONTAP permet des clones quasi instantanés et efficaces du stockage pour les migrations à chaud et à froid, et prend également en charge les copies quasi instantanées pour les migrations de volumes croisés de vVols. En raison de ces avantages considérables en matière d'efficacité du stockage, vous pouvez tirer pleinement parti des workloads vVols sous le "Garantie d'efficacité" programme. De même, si les clones multi-volumes à l'aide de VAAI ne répondent pas à vos besoins, vous serez probablement en mesure de relever vos défis business grâce aux améliorations apportées à l'expérience de copie des vVols.

Cas d'utilisation courants des vVols

Outre ces avantages, plusieurs cas d'utilisation courants sont également mentionnés ci-dessous pour le stockage vVol :

- **Provisionnement à la demande des machines virtuelles**
  - Cloud privé ou IaaS d'un Service Provider.
  - Exploitez l'automatisation et l'orchestration via la suite Aria (anciennement vRealize), OpenStack, etc.
- **Disques de première classe (FCDS)**
  - Volumes persistants VMware Tanzu Kubernetes Grid [TKG].
  - Proposez des services Amazon EBS avec une gestion indépendante du cycle de vie VMDK.
- **Approvisionnement à la demande des machines virtuelles temporaires**
  - Laboratoires de test et de développement
  - Environnements de formation

Bénéfices communs avec les vVols

Lorsqu'ils sont utilisés à leur plein avantage, comme dans les cas d'utilisation ci-dessus, les vVols apportent les améliorations spécifiques suivantes :

- La création de clones est rapide au sein d'un seul volume ou sur plusieurs volumes d'un cluster ONTAP. C'est un avantage par rapport aux clones classiques compatibles VAAI. Ils sont également efficaces en termes de stockage. Les clones d'un volume utilisent un clone de fichier ONTAP, qui ressemble aux volumes FlexClone et ne stockent que les modifications du fichier vVol source, de la LUN ou de l'espace de noms. Ainsi, les machines virtuelles à long terme pour la production ou d'autres applications sont créées rapidement, prennent un minimum d'espace et peuvent bénéficier de la protection au niveau des machines virtuelles (à l'aide du plug-in NetApp SnapCenter pour VMware vSphere, des snapshots gérés par VMware ou de la sauvegarde VADP) et de la gestion des performances (avec ONTAP QoS). Les clones intervolumes sont bien plus rapides avec les vVols qu'avec VAAI becuase avec VASA, nous pouvons créer le clone et y autoriser l'accès au niveau de la destination avant la fin de la copie. Les blocs de données sont copiés en arrière-plan pour remplir le vVol de destination. Cette approche est similaire à la méthode de déplacement de LUN sans interruption de ONTAP pour les LUN classiques.
- Les vVols sont la technologie de stockage idéale lors de l'utilisation de TKG avec vSphere CSI, fournissant des classes et des capacités de stockage distinctes gérées par l'administrateur vCenter.
- Les services de type Amazon EBS peuvent être fournis via les disques FCD, car un VMDK FCD, comme son nom l'indique, est citoyen de premier ordre dans vSphere et possède un cycle de vie qui peut être géré de manière indépendante, indépendamment des machines virtuelles auxquelles il peut être rattaché.

Liste de contrôle

Utilisez cette liste de contrôle d'installation pour garantir un déploiement réussi (mise à jour pour 10.3 et versions ultérieures).

1 **Planification initiale**

- Avant de commencer l'installation, vous devez vérifier "Matrice d'interopérabilité (IMT)" que votre déploiement a été certifié.
- Déterminez la taille et le type d'outils ONTAP nécessaires à la configuration de votre environnement. Pour plus d'informations, reportez-vous au "Limites de configuration pour le déploiement des outils ONTAP pour VMware vSphere".
- Déterminez si vous utiliserez des SVM mutualisés ou autorisez un accès complet au cluster. En cas d'utilisation de SVM mutualisés, vous devez disposer d'une LIF de gestion de SVM sur chaque SVM à utiliser. Cette LIF doit être accessible via le port 443 par les outils ONTAP.
- Déterminez si vous allez utiliser Fibre Channel (FC) pour la connectivité de stockage. Si c'est le cas, vous devez "configurer le zoning" sur vos commutateurs FC pour activer la connectivité entre les hôtes ESXi et les LIF FC du SVM.
- Déterminez si vous allez utiliser ONTAP Tools Storage Replication adapter (SRA) pour VMware site Recovery Manager (SRM) ou Live site Recovery (VLSR). Si tel est le cas, vous devrez accéder à l'interface de gestion du serveur SRM/VLSR pour installer SRA.
- Si vous utilisez la réplication SnapMirror gérée par les outils ONTAP (y compris, mais sans s'y limiter, la synchronisation active SnapMirror), votre administrateur ONTAP doit "Créez une relation entre clusters dans ONTAP" et "Créer une relation de pairs SVM intercluster dans ONTAP" avant de pouvoir utiliser les outils ONTAP avec SnapMirror.
- "Télécharger" Les outils ONTAP OVA et, si nécessaire, le fichier SRA tar.gz.

2 **Provisionnez les adresses IP et les enregistrements DNS**

- Demandez les informations IP suivantes à votre équipe réseau. Les trois premières adresses IP sont requises ; les nœuds deux et trois sont utilisés pour les déploiements scale-out de haute disponibilité (HA). Les enregistrements d'hôte DNS sont requis et tous les noms de nœud et toutes les adresses doivent se trouver sur le même VLAN et sous-réseau.
- Adresse de l'application des outils ONTAP _____ . _____ . _____ . _____
- Adresse des services internes _____ . _____ . _____ . _____
- Nom d'hôte DNS du nœud 1 _____________
- Adresse IP du nœud 1 _____ . _____ . _____ . _____
- Masque de sous-réseau _____ . ____ . \\ . _
- Passerelle par défaut _____ . _____ . _____ . _____
- Serveur DNS 1 _____ . _____ . _____ . _____
- Serveur DNS 2 _____ . _____ . _____ . _____
- Domaine de recherche DNS ______
- Nom d'hôte DNS du nœud deux (facultatif) __________
- Adresse IP du nœud deux (facultative) _____ . _____ . _____ . _____

- Nom d'hôte DNS du nœud 3 (facultatif) ___________
- Adresse IP du nœud 3 (facultative) _____ . _____ . _____ . _____
- Créez des enregistrements DNS pour toutes les adresses IP ci-dessus.

3 **Configuration du pare-feu réseau**

- Ouvrez les ports requis pour les adresses IP ci-dessus dans votre pare-feu réseau. Reportez-vous "Configuration requise pour les ports" à pour connaître la dernière mise à jour.

4 **Stockage**

- Un datastore sur un périphérique de stockage partagé est requis. Vous pouvez également utiliser une bibliothèque de contenu sur le même datastore que le nœud un pour faciliter le clonage rapide du modèle avec VAAI.
- Bibliothèque de contenu (requise uniquement pour HA) __________
- Nœud un magasin de données _____\
- Nœud deux datastore (facultatif, mais recommandé pour la haute disponibilité) ________\ _____
- Datastore du troisième nœud (facultatif, mais recommandé pour la haute disponibilité) __________\ __

5 **Déployez l'OVA**

- Notez que cette étape peut prendre jusqu'à 45 minutes
- "Déployez l'OVA" À l'aide du client vSphere.
- À l'étape 3 du déploiement d'OVA, sélectionnez l'option « personnaliser le matériel de cette machine virtuelle » et définissez les paramètres suivants à l'étape 10 :
- « Activer l'ajout à chaud du processeur »
- « Mémoire enfichable à chaud »

6 **Ajouter des vCenters aux outils ONTAP**

- "Ajouter des instances vCenter Server" Dans le Gestionnaire d'outils ONTAP.

7 **Ajout de systèmes back-end de stockage aux outils ONTAP**

- "Configurer les rôles et privilèges des utilisateurs ONTAP" Utilisation du fichier JSON inclus si vous n'utilisez pas admin.
- En cas d'utilisation en cas d'utilisation de la colocation du stockage avec des SVM
- "clusters intégrés" Dans le Gestionnaire d'outils ONTAP et les associer à des vCenters.
- "Ports SVM intégrés" Dans l'interface utilisateur vCenter des outils ONTAP.
- Si **NOT** utilise des SVM mutualisés
- "clusters intégrés" Directement dans l'interface utilisateur vCenter des outils ONTAP. Dans ce scénario, il

est également possible d'ajouter directement des SVM lorsque vous n'utilisez pas les vVols.

8 **Configuration des services de l'appliance (en option)**

- Pour utiliser vVols, vous devez d'abord "Modifiez les paramètres de l'appliance et activez le service VASA". En même temps, passez en revue les deux éléments suivants.
- Si vous prévoyez d'utiliser vVols en production, "haute disponibilité" avec les deux adresses IP facultatives ci-dessus.
- Si vous prévoyez d'utiliser ONTAP Tools Storage Replication adapter (SRA) pour VMware site Recovery Manager ou Live site Recovery, "Activation des services SRA".

9 **Certificats (facultatif)**

- Selon VMware, les certificats signés par une autorité de certification sont requis en cas d'utilisation de vVols avec plusieurs vCenters.
- Services Vasa __________
- Services administratifs _______

10 **Autres tâches post-déploiement**

- Créez des règles d'affinité pour les machines virtuelles dans un déploiement haute disponibilité.
- Si vous utilisez la haute disponibilité, Storage vMotion nœuds deux et trois vers des datastores séparés (facultatif, mais recommandé).
- "utilisez gérer les certificats" Dans le gestionnaire d'outils ONTAP pour installer les certificats signés par l'autorité de certification requis.
- Si vous avez activé SRA pour SRM/VLSR pour protéger les datastores traditionnels, "Configurez SRA sur l'appliance VMware Live site Recovery".
- Configurer les sauvegardes natives pour "RPO proche de zéro".
- Configurer des sauvegardes régulières sur d'autres supports de stockage.

Utilisation de vVols avec ONTAP

La clé de l'utilisation de vVols avec NetApp est l'outil ONTAP pour VMware vSphere, qui fait office d'interface VASA (vSphere API for Storage Awareness) Provider pour les systèmes ONTAP 9 de NetApp.

Les outils ONTAP incluent également les extensions d'interface utilisateur vCenter, les services d'API REST, les adaptateurs de réplication du stockage pour VMware site Recovery Manager / Live site Recovery, les outils de surveillance et de configuration d'hôte, ainsi qu'une série de rapports qui vous aident à mieux gérer votre environnement VMware.

Produits et documentation

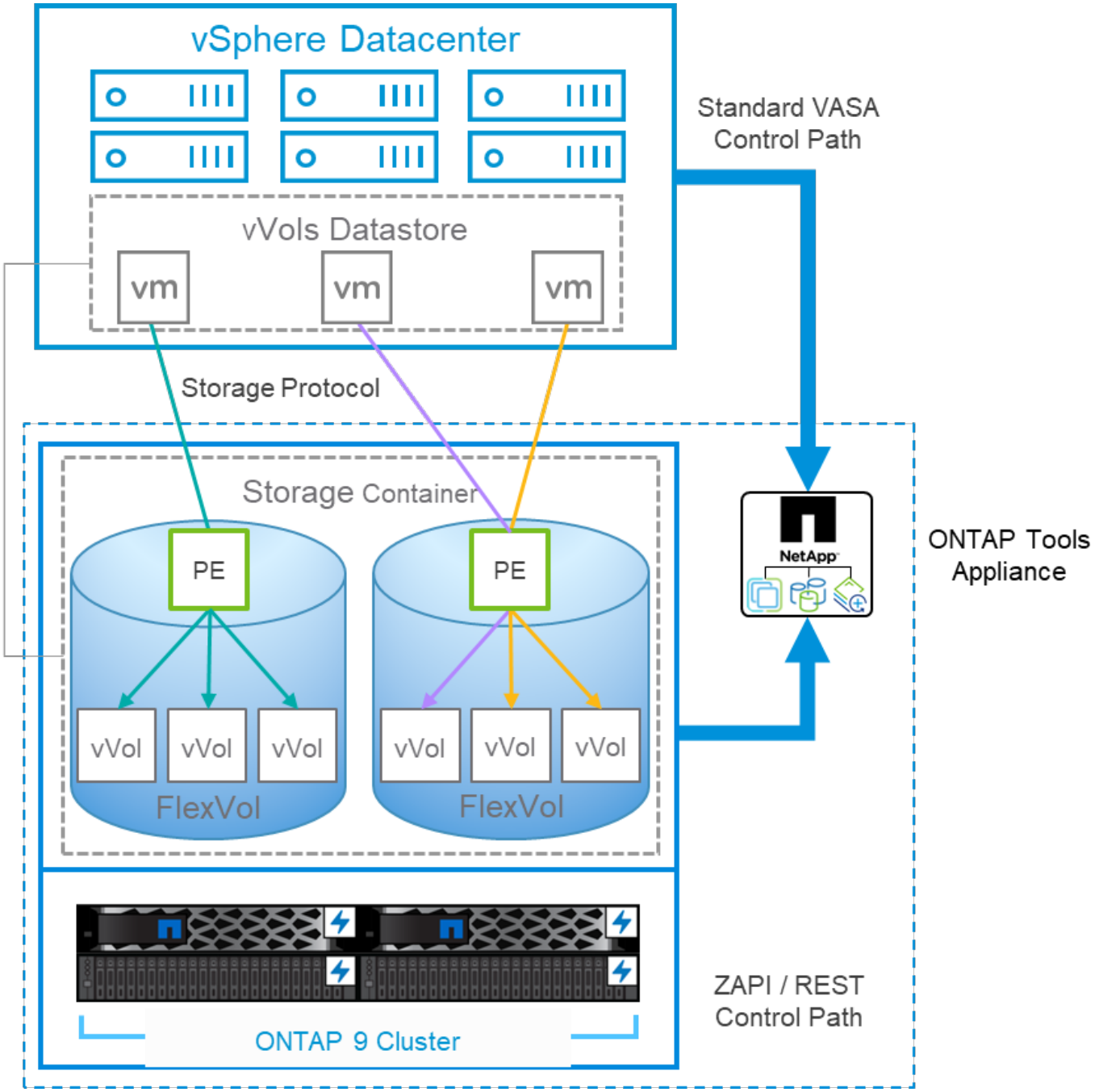
La licence ONTAP One inclut toutes les licences nécessaires pour utiliser les vVols avec les systèmes ONTAP. La seule autre exigence est les outils ONTAP gratuits OVA, qui agit en tant que fournisseur VASA. Dans un environnement vVols, le logiciel VASA Provider traduit les fonctionnalités des baies en attributs basés sur des règles qui peuvent être exploitées via les API VASA sans que l'administrateur vSphere n'ait à savoir comment

les fonctionnalités sont gérées en arrière-plan. Cela permet une consommation dynamique de la capacité de stockage allouée sur la base de règles, ce qui évite de créer manuellement des datastores classiques et de gérer leurs taux de consommation de stockage individuels. En bref, les vVols simplifient la gestion du stockage d'entreprise et l'affranchit de l'administrateur vSphere pour qu'il puisse se concentrer sur la couche de virtualisation.

Pour les clients qui utilisent VMware Cloud Foundation avec VSAN, les vVols peuvent être ajoutés à n'importe quel domaine de gestion ou de workload en tant que stockage supplémentaire. VVols s'intègre de façon transparente avec VSAN via un framework de gestion commun basé sur des règles de stockage.

La gamme d'outils ONTAP 10 nouvelle génération modernise les fonctionnalités précédentes grâce à une architecture évolutive, conteneurisée et basée sur des microservices qui peut être déployée via une simple appliance de format OVA sur ESXi. Les outils ONTAP 10 combinent toutes les fonctionnalités de trois anciens dispositifs et produits dans un déploiement unique. Pour la gestion des vVols, vous utiliserez les extensions intuitives de l'interface utilisateur vCenter ou les API REST pour les outils ONTAP VASA Provider. Notez que le composant SRA est destiné aux datastores classiques ; VMware site Recovery Manager n'utilise pas SRA pour les vVols.

ONTAP Tools VASA Provider architecture lors de l'utilisation d'iSCSI ou FCP avec des systèmes unifiés



Installation du produit

Pour les nouvelles installations, déployez l'appliance virtuelle dans votre environnement vSphere. Une fois déployé, vous pouvez vous connecter à l'interface utilisateur du gestionnaire ou utiliser les API REST pour faire évoluer votre déploiement verticalement ou horizontalement, intégrer les vCenters (qui enregistrent le plug-in avec vCenter), intégrer les systèmes de stockage et associer les systèmes de stockage à vos vCenters. L'intégration de systèmes de stockage dans l'interface du gestionnaire d'outils ONTAP et l'association de clusters à des vCenters sont uniquement nécessaires si vous prévoyez d'utiliser la colocation sécurisée avec des SVM dédiés. Sinon, vous pouvez simplement intégrer le ou les clusters de stockage souhaités dans les extensions de l'interface utilisateur vCenter des outils ONTAP ou à l'aide des API REST.

Reportez-vous à "Déploiement du stockage vVols" dans ce document, ou "Documentation sur les outils ONTAP pour VMware vSphere".

> Il est recommandé de stocker vos outils ONTAP et appliances vCenter sur des datastores NFS ou VMFS classiques afin d'éviter tout conflit d'interdépendance. Étant donné que les outils vCenter et ONTAP doivent communiquer entre eux lors des opérations vVols, n'installez pas et ne déplacez pas les appliances ONTAP Tools ou vCenter Server (VCSA) vers le stockage vVols qu'ils gèrent. Dans ce cas, le redémarrage de l'appliance vCenter ou des outils ONTAP peut entraîner une interruption de l'accès au plan de contrôle et une incapacité de l'appliance à démarrer.

Les mises à niveau des outils ONTAP sans déplacement des données sont prises en charge grâce au fichier ISO de mise à niveau disponible en téléchargement "Outils ONTAP pour VMware vSphere 10 - Téléchargements" sur le site du support NetApp (connexion requise). Suivez les "Mise à niveau des outils ONTAP pour VMware vSphere 10.x vers la version 10.3" instructions du guide pour mettre à niveau l'appareil. Il est également possible d'effectuer une mise à niveau côte à côte des outils ONTAP 9.13 à 10.3. Reportez-vous "Migrez des outils ONTAP pour VMware vSphere 9.x vers la version 10.3" à la pour plus d'informations sur ce sujet.

Pour le dimensionnement de votre appliance virtuelle et la compréhension des limites de configuration, reportez-vous à la section "Limites de configuration pour le déploiement des outils ONTAP pour VMware vSphere"

Documentation produit

La documentation suivante est disponible pour vous aider à déployer les outils ONTAP.

"Documentation sur les outils ONTAP pour VMware vSphere"

Commencez

- "Notes de mise à jour"
- "Présentation des outils ONTAP pour VMware vSphere"
- "Déployez les outils ONTAP"
- "Mettez à niveau les outils ONTAP"

Utilisez les outils ONTAP

- "Provisionner les datastores"
- "Configurez le contrôle d'accès basé sur des rôles"
- "Configurez la haute disponibilité"
- "Modifier les paramètres de l'hôte VMware ESXi"

Protéger et gérer les datastores

- "Configurez vSphere Metro Storage Cluster (vMSC) à l'aide des outils ONTAP et de la synchronisation active SnapMirror"
- "Protection des machines virtuelles" Avec SRM
- "Surveillez les clusters, les datastores et les machines virtuelles"

Tableau de bord VASA Provider

Le fournisseur VASA inclut un tableau de bord contenant des informations sur les performances et la capacité

des VM vVols individuelles. Ces informations proviennent directement de ONTAP pour les fichiers et les LUN VVol, notamment la latence, les IOPS, le débit, etc. Il est activé par défaut lors de l'utilisation de toutes les versions de ONTAP 9 actuellement prises en charge. Notez qu'après la configuration initiale, le tableau de bord peut contenir jusqu'à 30 minutes de données.

Autres pratiques exemplaires

L'utilisation des vVols de ONTAP avec vSphere est simple et suit les méthodes vSphere publiées (consultez la documentation utilisation des volumes virtuels sous vSphere Storage in VMware pour votre version d'ESXi). Voici quelques autres pratiques à prendre en compte avec ONTAP.

Limites

En général, ONTAP supporte les limites vVols définies par VMware (voir publié "Configuration maximale"). Vérifiez toujours les limites mises à jour du "NetApp Hardware Universe" nombre et de la taille des LUN, des espaces de noms et des fichiers.

Utilisez les outils ONTAP pour les extensions d'interface utilisateur ou les API REST de VMware vSphere pour provisionner les datastores vVols et les terminaux de protocole.

Même s'il est possible de créer des datastores vVols avec l'interface vSphere générale, l'utilisation des outils ONTAP crée automatiquement des terminaux de protocole selon les besoins et crée des volumes FlexVol (non requis avec ASA r2) en utilisant les bonnes pratiques ONTAP. Il vous suffit de cliquer avec le bouton droit de la souris sur l'hôte/le cluster/le data Center, puis de sélectionner *ONTAP Tools* et *provisioning datastore*. Ensuite, il vous suffit de choisir les options vVols souhaitées dans l'assistant.

Ne stockez jamais l'appliance ONTAP Tools ou l'appliance vCenter Server (VCSA) sur un datastore vVols qu'ils gèrent.

Cela peut entraîner une « situation de poulet et d'œuf » si vous devez redémarrer les appareils car ils ne pourront pas réassocier leurs propres vVols pendant qu'ils redémarrent. Vous pouvez les stocker sur un datastore vVols géré par un autre outil ONTAP et un déploiement vCenter.

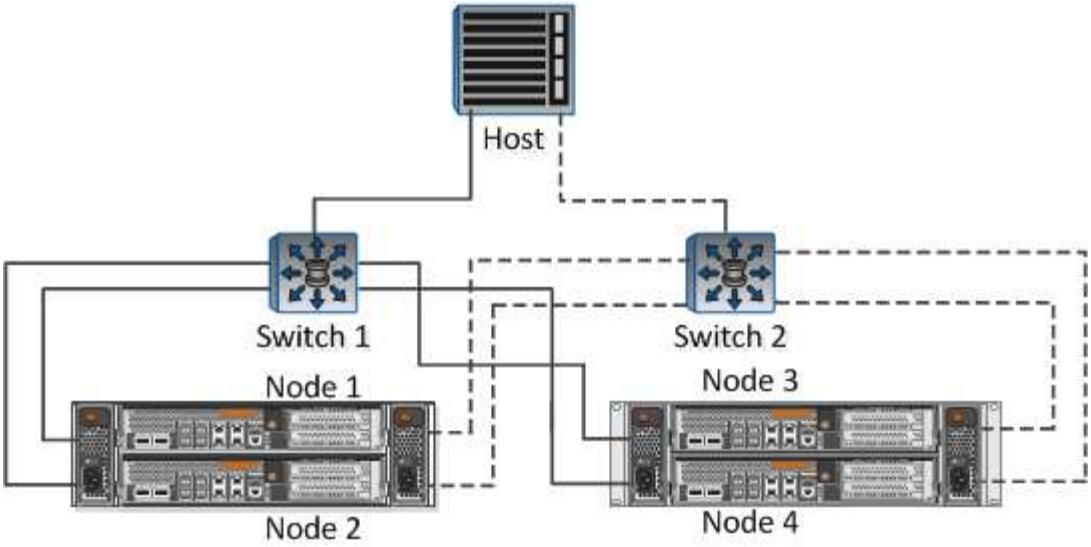
Évitez les opérations vVols sur différentes versions de ONTAP.

Les fonctionnalités de stockage prises en charge telles que la QoS, le personnalité et bien d'autres encore ont changé dans plusieurs versions du fournisseur VASA, et certaines dépendent de la version de ONTAP. L'utilisation de différentes versions dans un cluster ONTAP ou le déplacement de vVols entre clusters avec différentes versions peut entraîner un comportement inattendu ou des alarmes de conformité.

Zone votre fabric Fibre Channel avant d'utiliser FCP pour vVols.

Le fournisseur VASA des outils ONTAP se charge de la gestion des igroups FCP et iSCSI ainsi que des sous-systèmes NVMe dans ONTAP en fonction des initiateurs détectés d'hôtes ESXi gérés. Toutefois, il ne s'intègre pas aux commutateurs Fibre Channel pour gérer la segmentation. La segmentation doit être effectuée conformément aux meilleures pratiques avant tout provisionnement. Voici un exemple de segmentation à un seul initiateur sur quatre systèmes ONTAP :

Segmentation à un seul initiateur :



Pour plus d'informations sur les meilleures pratiques, reportez-vous aux documents suivants :

"*TR-4080 meilleures pratiques pour le SAN moderne ONTAP 9*"

"*TR-4684 implémentation et configuration de SAN modernes avec NVMe-of*"

Planifier vos volumes FlexVol de support en fonction de vos besoins.

Pour les systèmes non ASA r2, il peut être souhaitable d'ajouter plusieurs volumes de sauvegarde à votre datastore vVols pour répartir la charge de travail sur le cluster ONTAP, pour prendre en charge différentes options de règles ou pour augmenter le nombre de LUN ou de fichiers autorisés. Toutefois, si vous avez besoin d'une efficacité de stockage maximale, placez l'ensemble de vos volumes en arrière-forme sur un seul agrégat. Si des performances de clonage maximales sont requises, envisagez d'utiliser un seul volume FlexVol et de conserver vos modèles ou votre bibliothèque de contenu dans le même volume. Le fournisseur VASA délègue de nombreuses opérations de stockage vVols à ONTAP, notamment la migration, le clonage et les copies Snapshot. Cette opération est réalisée au sein d'un seul volume FlexVol, ce qui permet d'utiliser des clones de fichiers peu encombrants et de les mettre presque instantanément à disposition. Sur des volumes FlexVol, les copies sont rapidement disponibles et utilisent la déduplication et la compression à la volée. Toutefois, l'efficacité du stockage maximale ne peut pas être restaurée tant que des tâches en arrière-plan ne sont pas exécutées sur des volumes utilisant la déduplication et la compression en arrière-plan. Selon la source et la destination, une certaine efficacité peut être dégradée.

Avec les systèmes ASA r2, cette complexité n'est plus liée à l'abstraction du concept de volume ou d'agrégat par rapport à l'utilisateur. Le placement dynamique est géré automatiquement et des terminaux de protocole sont créés en fonction des besoins. Des terminaux supplémentaires peuvent être créés automatiquement à la volée si une évolutivité supplémentaire est nécessaire.

Pensez à utiliser Max IOPS pour contrôler des machines virtuelles inconnues ou tester des machines virtuelles.

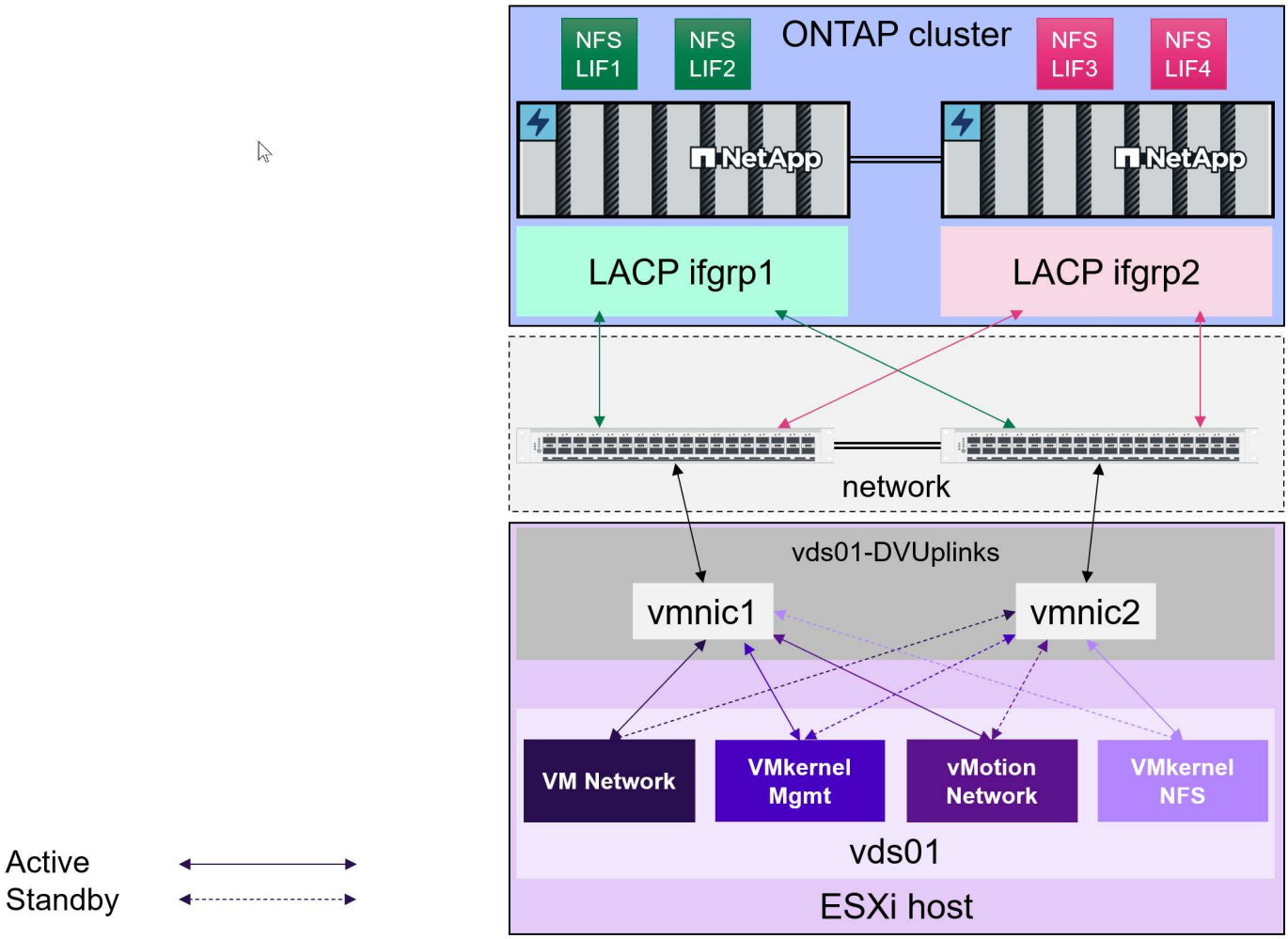
Disponible pour la première fois dans VASA Provider 7.1, Max IOPS peut être utilisé pour limiter les IOPS à un vVol spécifique pour une charge de travail inconnue afin d'éviter tout impact sur d'autres charges de travail plus stratégiques. Pour plus d'informations sur la gestion des performances, consultez le Tableau 4.

Assurez-vous d'avoir suffisamment de LIFs de données. Reportez-vous à la "Déploiement du stockage vVols".

Suivre toutes les meilleures pratiques du protocole.

Reportez-vous aux autres guides des meilleures pratiques de NetApp et VMware spécifiques au protocole sélectionné. En général, il n'y a pas d'autres changements que ceux déjà mentionnés.

Exemple de configuration réseau utilisant vVols sur NFS v3



Déploiement de vVols sur les systèmes AFF, ASA, ASA r2 et FAS

Suivez ces bonnes pratiques pour créer du stockage vVols pour vos machines virtuelles.

Le provisionnement des datastores vVols implique plusieurs étapes. Les systèmes ASA r2 de NetApp sont conçus pour les charges de travail VMware et offrent une expérience utilisateur différente des systèmes ONTAP classiques. Lors de l'utilisation de systèmes ASA r2, la configuration des outils ONTAP versions 10.3 ou ultérieures requiert moins d'étapes, en incluant les extensions d'interface utilisateur et la prise en charge de l'API REST optimisées pour la nouvelle architecture de stockage.

Préparation à la création de datastores vVols avec les outils ONTAP

Vous pouvez ignorer les deux premières étapes du processus de déploiement si vous utilisez déjà les outils ONTAP pour gérer, automatiser et générer des rapports sur votre stockage VMFS ou NFS classique existant. Vous pouvez également vous référer à ce document complet "liste de contrôle" pour le déploiement et la configuration des outils ONTAP.

1. Créer la machine virtuelle de stockage (SVM) et sa configuration de protocole. Notez que cela n'est pas forcément nécessaire pour les systèmes ASA r2, car ils disposent généralement d'un SVM unique pour les

services de données. Vous sélectionnerez NVMe/FC (outils ONTAP 9.13 uniquement), NFS v3, NFS v4.1, iSCSI, FCP ou un mélange de ces options. Les protocoles NVMe/TCP et NVMe/FC peuvent également être utilisés pour les datastores VMFS classiques avec les outils ONTAP version 10.3 et ultérieures. Vous pouvez utiliser les assistants ONTAP System Manager ou la ligne de commande du cluster shell.

- "Attribuez des tiers locaux (agrégats) aux SVM" Pour tous les systèmes non ASA r2.
- Au moins une LIF par nœud pour chaque connexion switch/fabric. Il est recommandé de créer au moins deux par nœud pour les protocoles FCP, iSCSI ou NVMe. Une LIF par nœud est suffisante pour les vVols basés sur NFS, mais cette LIF doit être protégée par un ifgroup LACP. Reportez-vous aux sections "Configurer la présentation des LIFs" et "Combinaison de ports physiques pour créer des groupes d'interfaces" pour plus de détails.
- Si vous avez l'intention d'utiliser les identifiants SVM scoped pour vos vCenters locataires, au moins une LIF de gestion par SVM.
- Si vous prévoyez d'utiliser SnapMirror, assurez-vous que votre source et votre cible "Les clusters ONTAP et les SVM sont peering".
- Pour les systèmes non ASA r2, des volumes peuvent être créés à ce stade, mais il est recommandé de laisser l'assistant *provisioning datastore* dans les outils ONTAP les créer. La seule exception à cette règle est que vous prévoyez d'utiliser la réplication vVols avec VMware site Recovery Manager et les outils ONTAP 9,13. Cette configuration est plus simple avec des volumes FlexVol préexistants avec des relations SnapMirror existantes. N'oubliez pas d'activer la QoS sur les volumes à utiliser pour les vVols, car ceux-ci doivent être gérés par les outils SPBM et ONTAP.

2. "Déployez les outils ONTAP pour VMware vSphere" Utilisation du fichier OVA téléchargé depuis le site de support NetApp.
   - Les outils ONTAP 10.0 et versions ultérieures prennent en charge plusieurs serveurs vCenter par appliance. Il n'est plus nécessaire de déployer une appliance d'outils ONTAP par serveur vCenter.
     - Si vous prévoyez de connecter plusieurs vCenters à une seule instance d'outils ONTAP, vous devez créer et installer des certificats signés par une autorité de certification. Reportez-vous "Gérer les certificats" à pour les étapes.
   - Depuis 10.3, les outils ONTAP se déploient désormais en tant qu'appliance petit type à nœud unique, adaptée à la plupart des workloads non vVols.

> - Il est recommandé "Outils ONTAP scale-out" de passer à la version 10.3 et ultérieures de la configuration haute disponibilité (HA) à 3 nœuds pour toutes les charges de travail de production. Dans les laboratoires ou à des fins de test, il est possible d'utiliser un déploiement à un seul nœud.
> - La meilleure pratique recommandée pour l'utilisation des vVols de production est d'éliminer tout point de défaillance unique. Créez des règles d'anti-affinité pour empêcher les machines virtuelles des outils ONTAP de s'exécuter ensemble sur le même hôte. Après le déploiement initial, il est également recommandé d'utiliser Storage vMotion pour placer les machines virtuelles des outils ONTAP dans différents datastores. En savoir plus sur "Utilisation des règles d'affinité sans vSphere DRS" ou "Créez une règle d'affinité VM-VM". Vous devez également planifier des sauvegardes fréquentes, et/ou "utilisez l'utilitaire de sauvegarde de configuration intégré".

1. Configurez les outils ONTAP 10.3 en fonction de votre environnement.
   - "Ajouter des instances vCenter Server" Dans l'interface utilisateur du gestionnaire d'outils ONTAP.
   - Les outils ONTAP 10.3 prennent en charge la colocation sécurisée. Si vous n'avez pas besoin d'une colocation sécurisée, il vous suffit d' "Ajoutez vos clusters ONTAP"accéder au menu des outils ONTAP dans vCenter, de cliquer sur *systèmes back-end* et de cliquer sur le bouton *ajouter*.

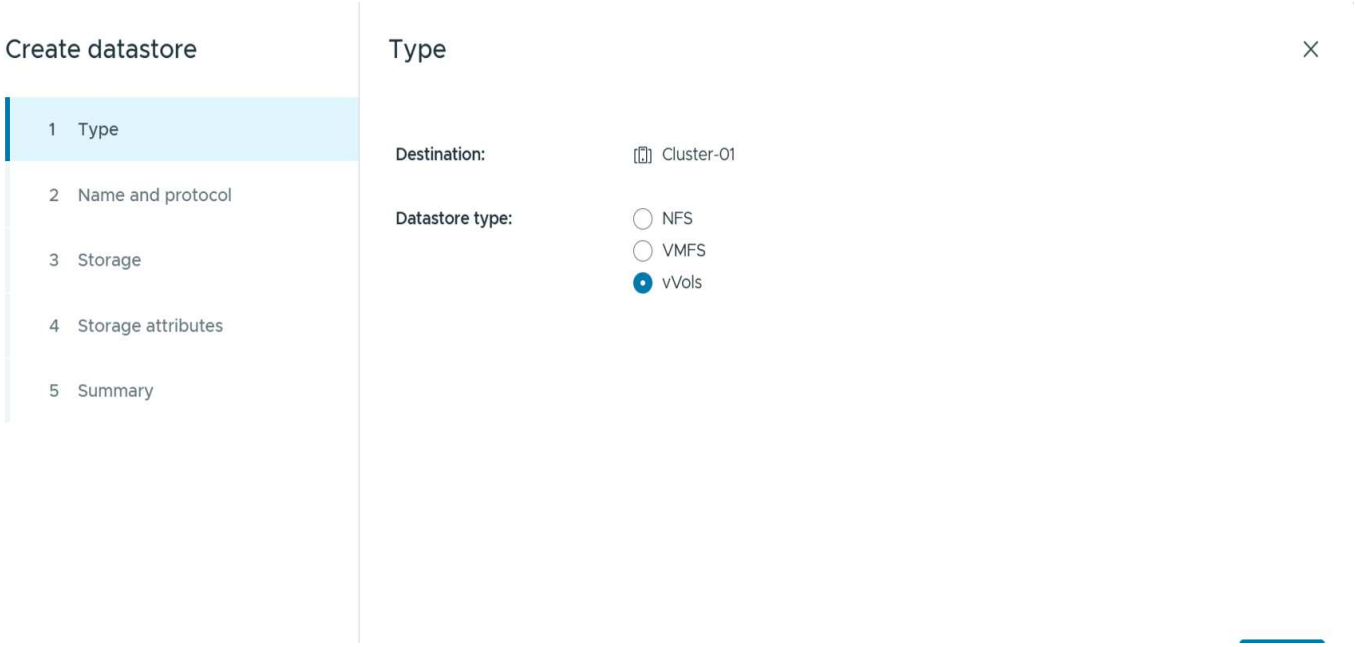
- Dans un environnement mutualisé sécurisé où vous souhaitez déléguer des SVM spécifiques à des vCenters spécifiques, vous devez procéder comme suit.
  - Connectez-vous à l'interface du gestionnaire d'outils ONTAP
  - "Intégration du cluster de stockage"
  - "Associer un back-end de stockage à une instance vCenter Server"
  - Fournir les informations d'identification de SVM spécifiques à l'administrateur vCenter qui ajoutera ensuite la SVM en tant que back-end de stockage dans le menu ONTAP Tools Storage backend de vCenter.

- Il est recommandé de créer des rôles RBAC pour vos comptes de stockage.
- Les outils ONTAP incluent un fichier JSON contenant les autorisations de rôle requises par les comptes d'accès au stockage des outils ONTAP. Vous pouvez télécharger le fichier JSON vers ONTAP System Manager pour simplifier la création des rôles et utilisateurs RBAC.
- Pour en savoir plus sur les rôles RBAC de ONTAP, rendez-vous sur "Configurer les rôles et privilèges des utilisateurs ONTAP".

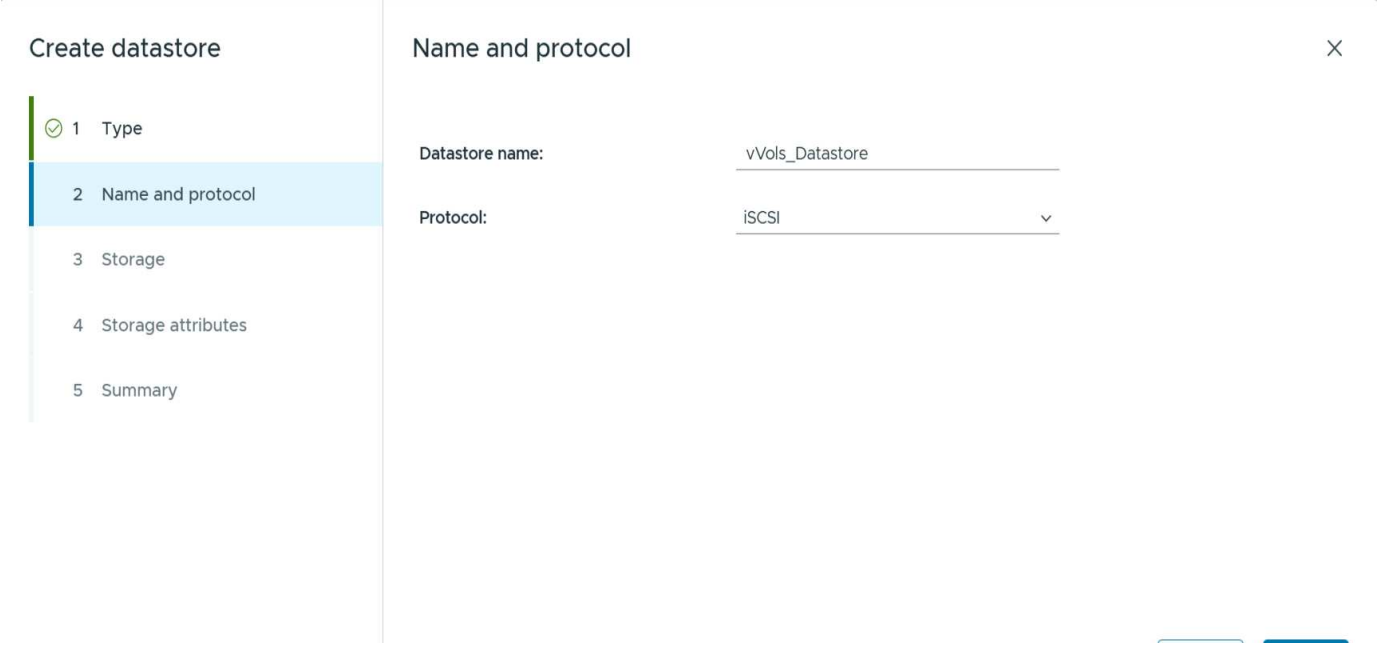
Parce que de nombreuses API utilisées pour les vVols ne sont disponibles qu'au niveau du cluster, l'intégralité du cluster doit être intégrée dans l'interface utilisateur du gestionnaire des outils ONTAP.

Création de datastores vVols avec les outils ONTAP

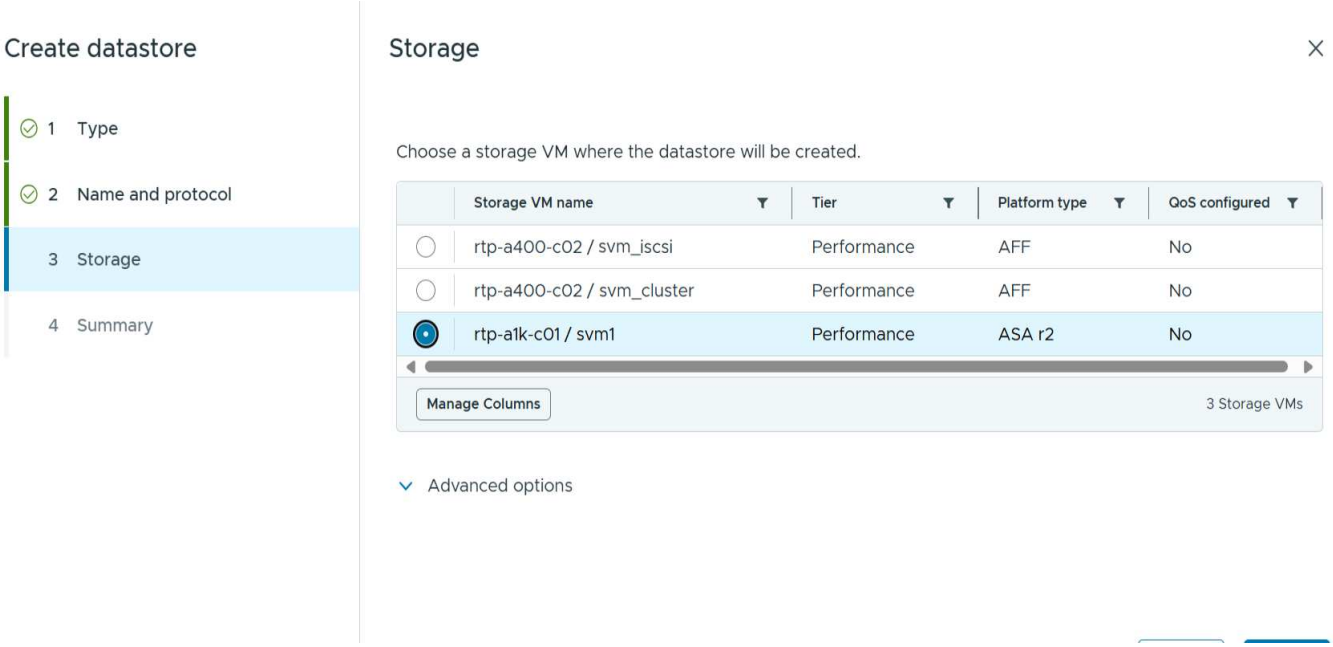
Cliquez avec le bouton droit de la souris sur l'hôte, le cluster ou le data Center sur lequel vous souhaitez créer le datastore vVols, puis sélectionnez *ONTAP Tools > Provision datastore*.



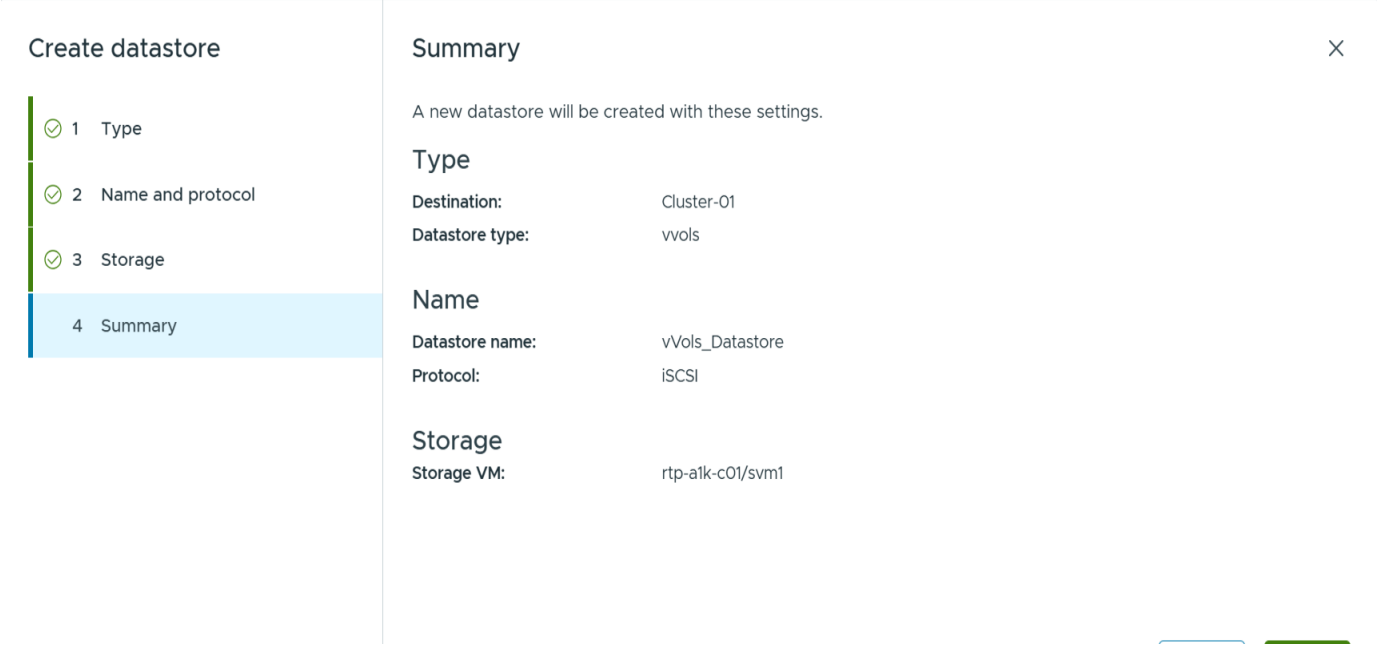
- Choisissez vVols, indiquez un nom significatif et sélectionnez le protocole souhaité. Vous pouvez également fournir une description du datastore.
  - Outils ONTAP 10.3 avec ASA r2.



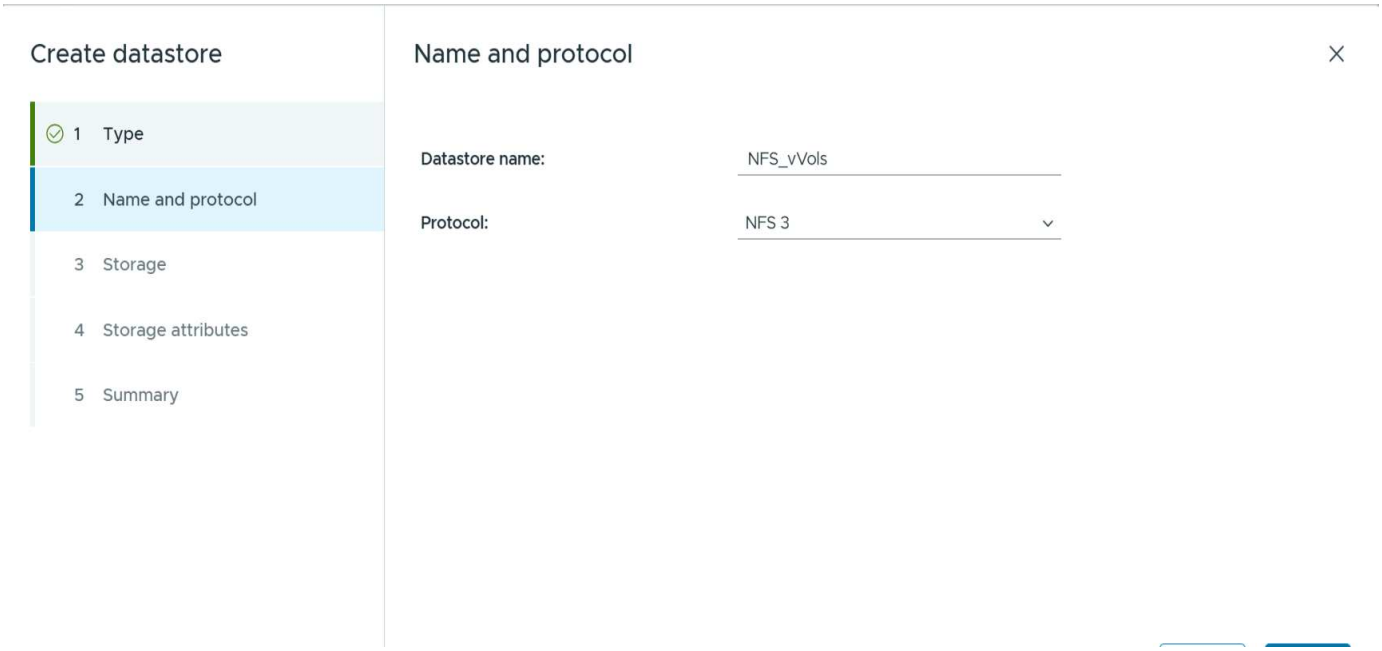
- Sélectionner le SVM du système ASA r2 et cliquer sur *Next*.



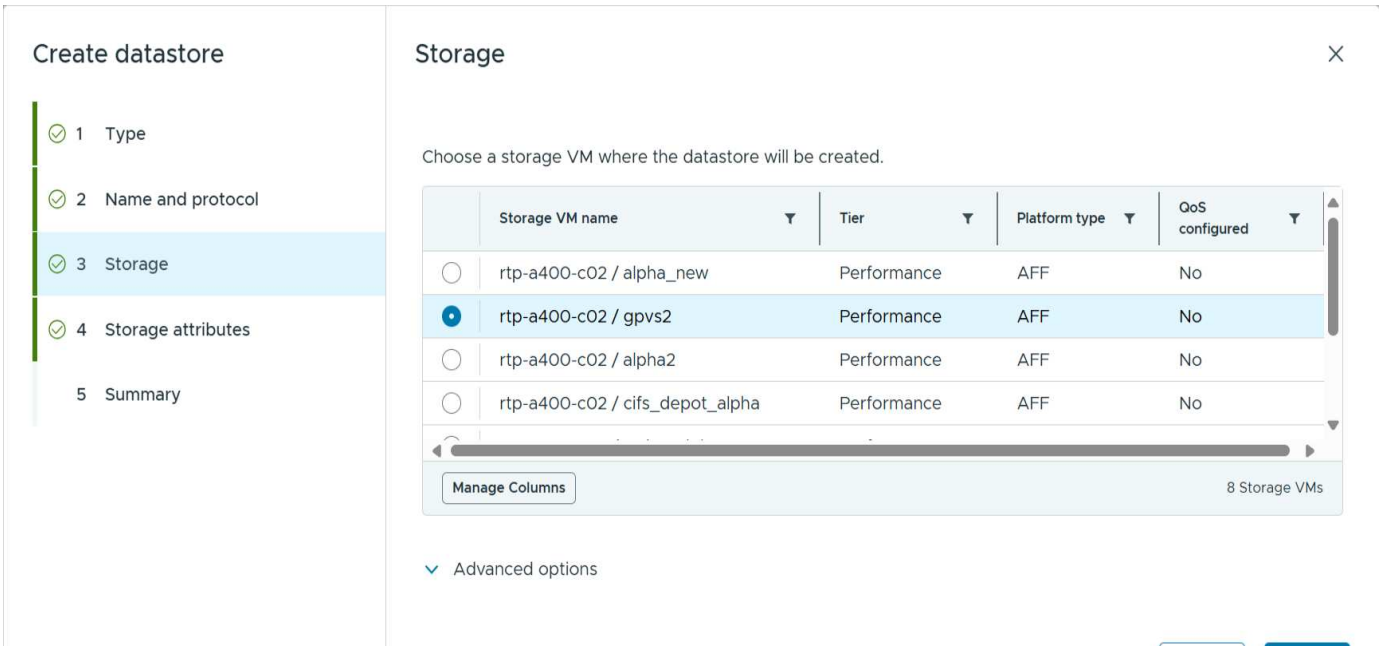
- Cliquez sur *Finish*



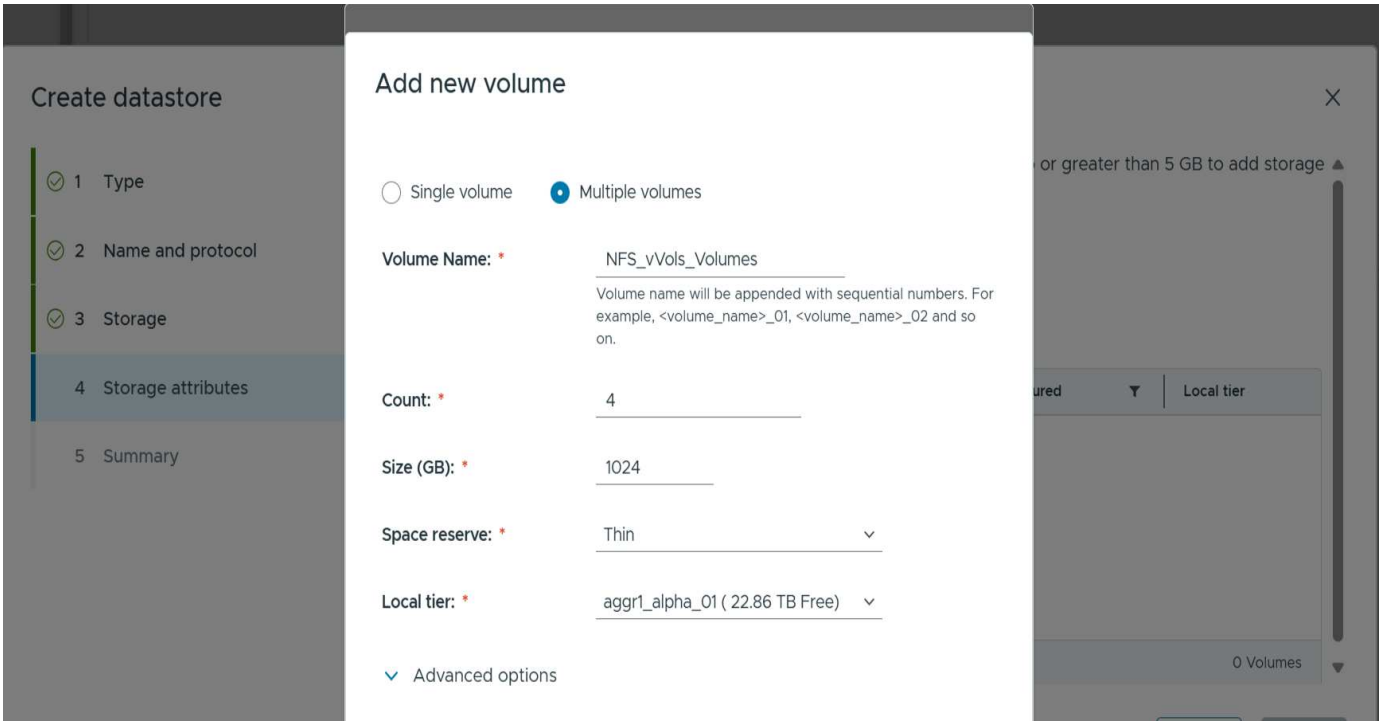
- C’est aussi simple que cela !
  - Outils ONTAP 10.3 avec ONTAP FAS, AFF et ASA antérieurs à ASA r2.
- Sélectionnez le protocole

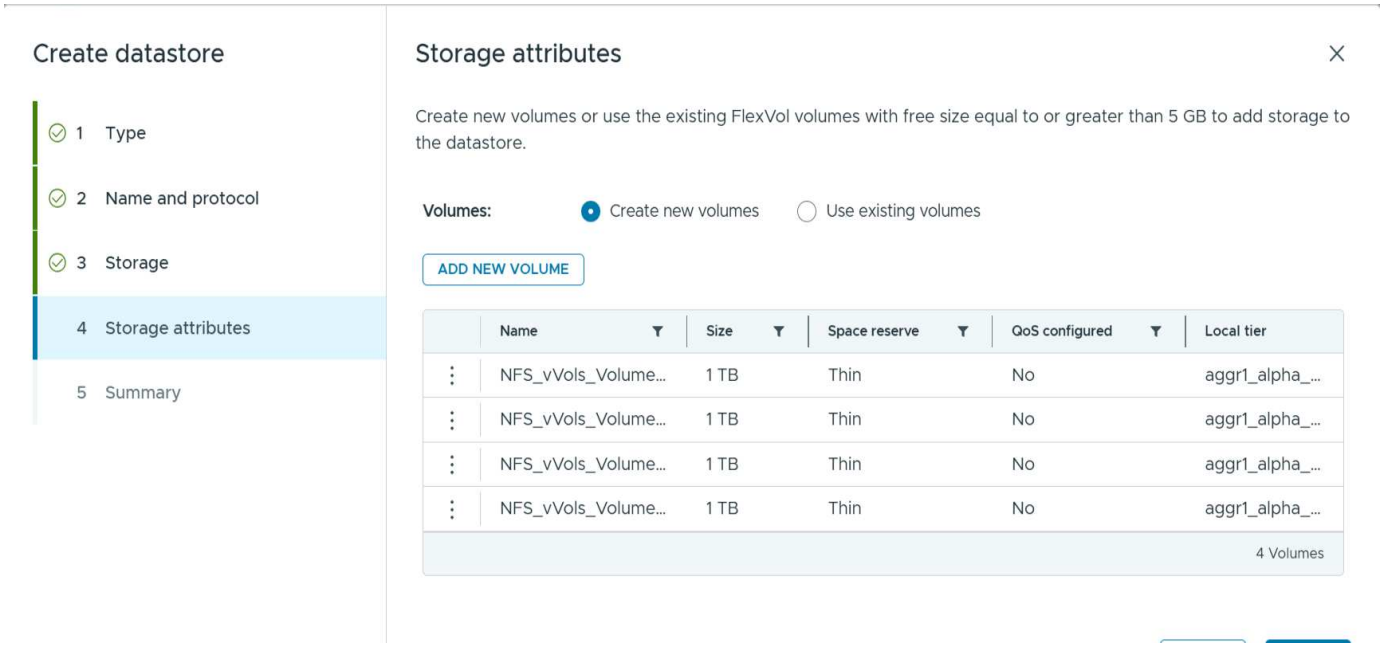


- Sélectionner le SVM et cliquer sur *Next*.

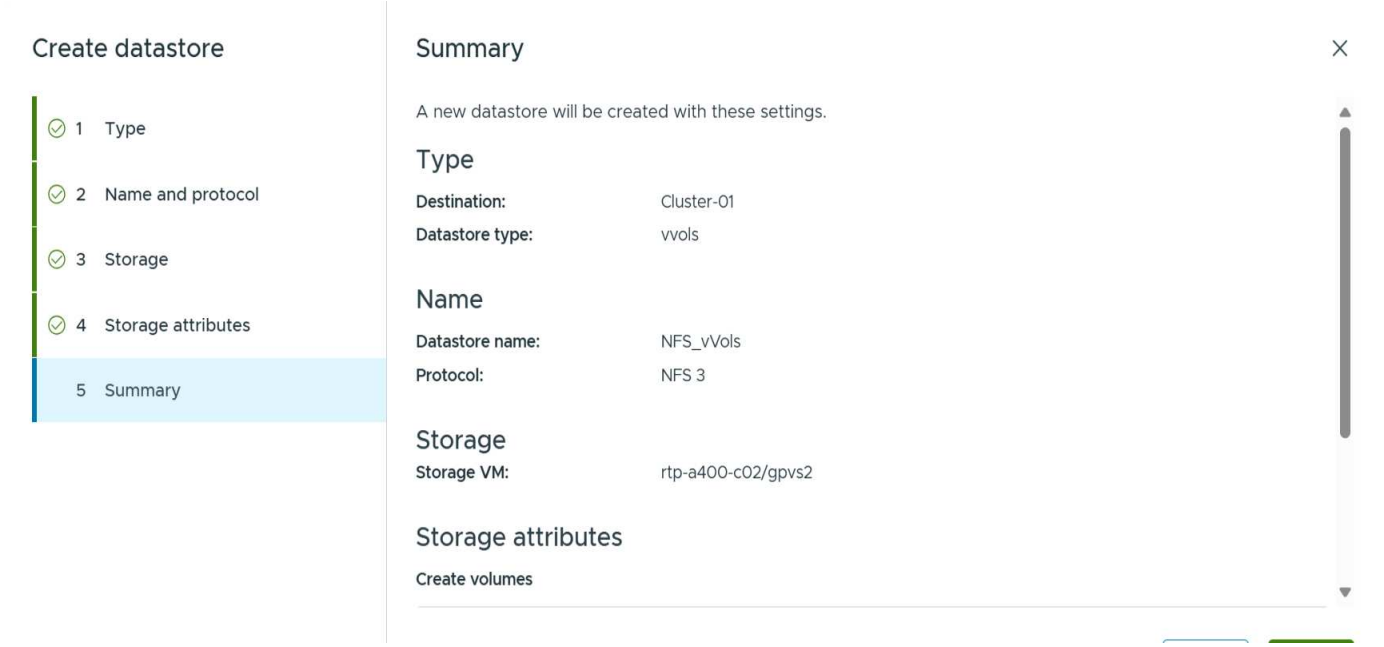


- Cliquez sur *add New volumes* ou *use existing volume* et spécifiez les attributs. Notez que dans les outils ONTAP 10.3 vous pouvez demander la création simultanée de plusieurs volumes. Vous pouvez également ajouter manuellement plusieurs volumes pour les équilibrer dans le cluster ONTAP. Cliquez sur *Next*

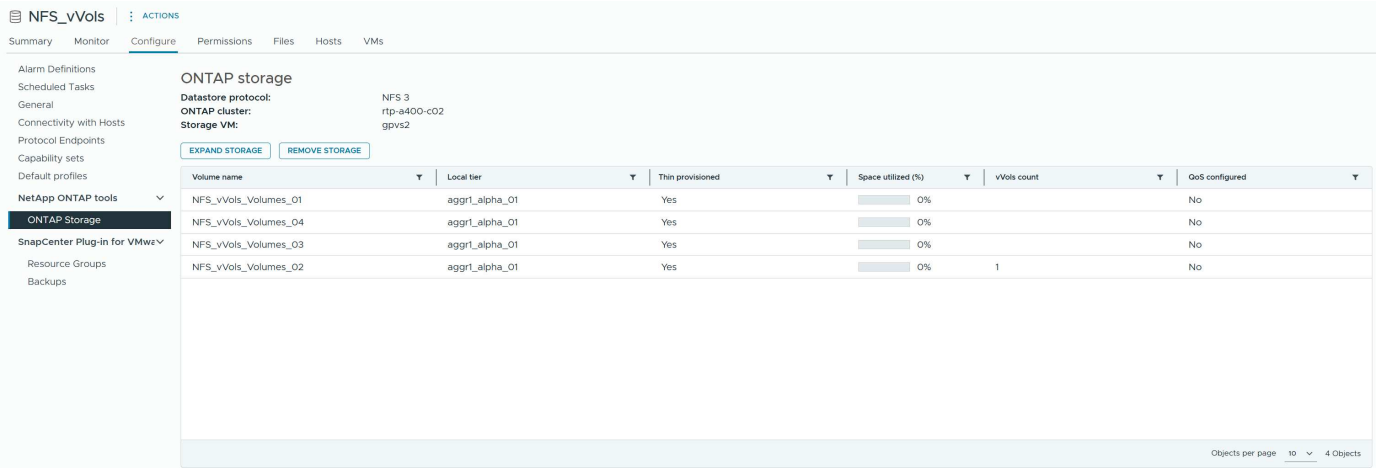




- Cliquez sur *Finish*



- Les volumes affectés s’affichent dans le menu Outils ONTAP de l’onglet configurer du datastore.



- Vous pouvez désormais créer des stratégies de stockage de machine virtuelle à partir du menu *Policies and Profiles* de l'interface utilisateur vCenter.

Migration des machines virtuelles des datastores classiques vers des vVols

La migration des machines virtuelles des datastores traditionnels vers un datastore vVols est aussi simple que le déplacement de machines virtuelles entre des datastores traditionnels. Il vous suffit de sélectionner la ou les machines virtuelles, puis de sélectionner migrer dans la liste actions et de sélectionner un type de migration de *modifier le stockage uniquement*. Lorsque vous y êtes invité, sélectionnez une règle de stockage de machine virtuelle correspondant à votre datastore vVols. Les opérations de copie de migration peuvent être déchargées à l'aide de vSphere 6.0 et versions ultérieures pour les migrations de SAN VMFS vers des vVols, mais pas des VMDK NAS vers des vVols.

Gestion des machines virtuelles avec des règles

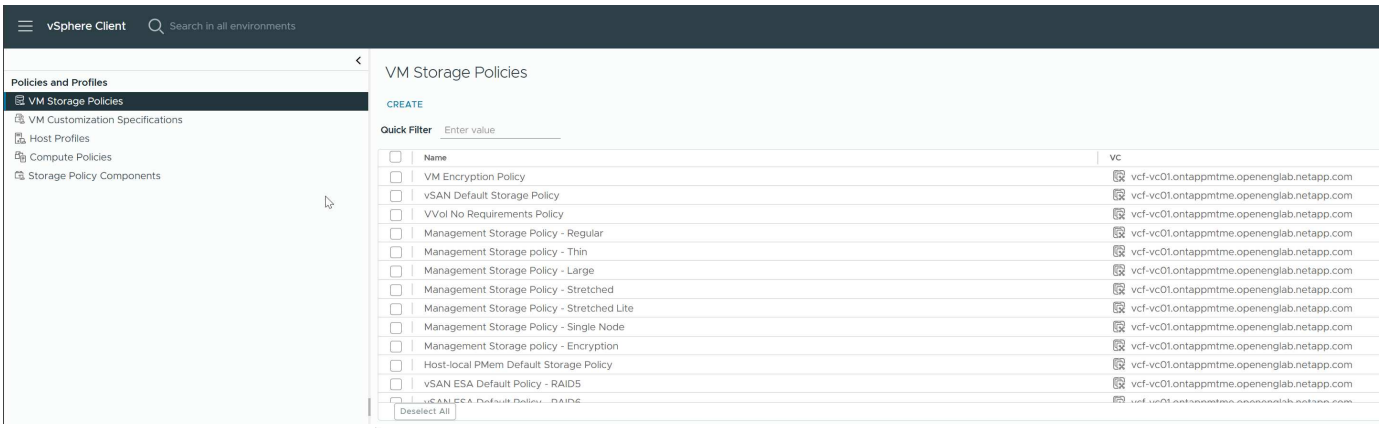
Pour automatiser le provisionnement du stockage avec la gestion basée sur des règles, vous devez créer des règles de stockage de machines virtuelles mappées sur les capacités de stockage souhaitées.

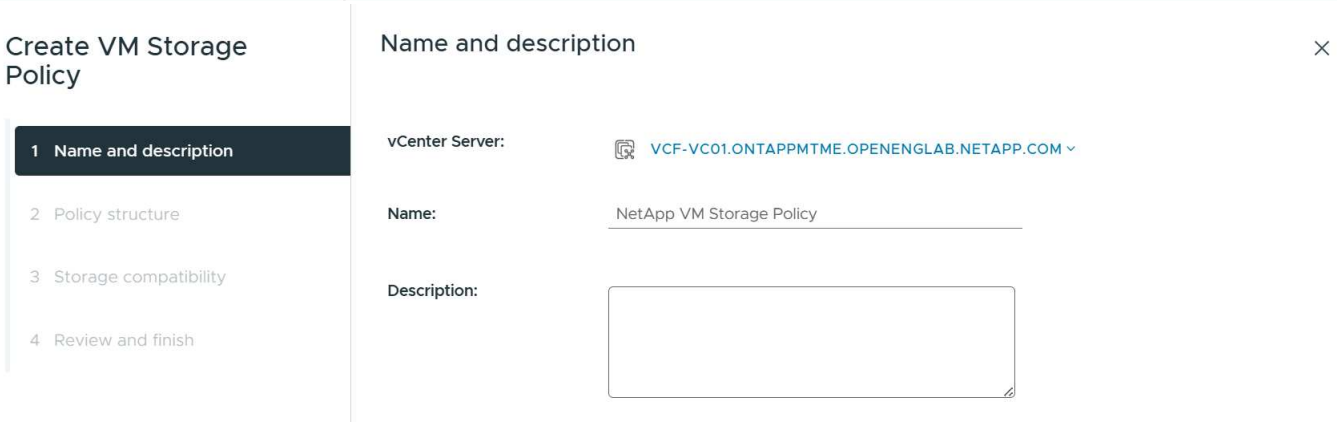
> Les outils ONTAP 10.0 et versions ultérieures n'utilisent plus les profils de capacité de stockage comme les versions précédentes. Au contraire, les fonctionnalités de stockage sont directement définies dans la stratégie de stockage de la machine virtuelle.

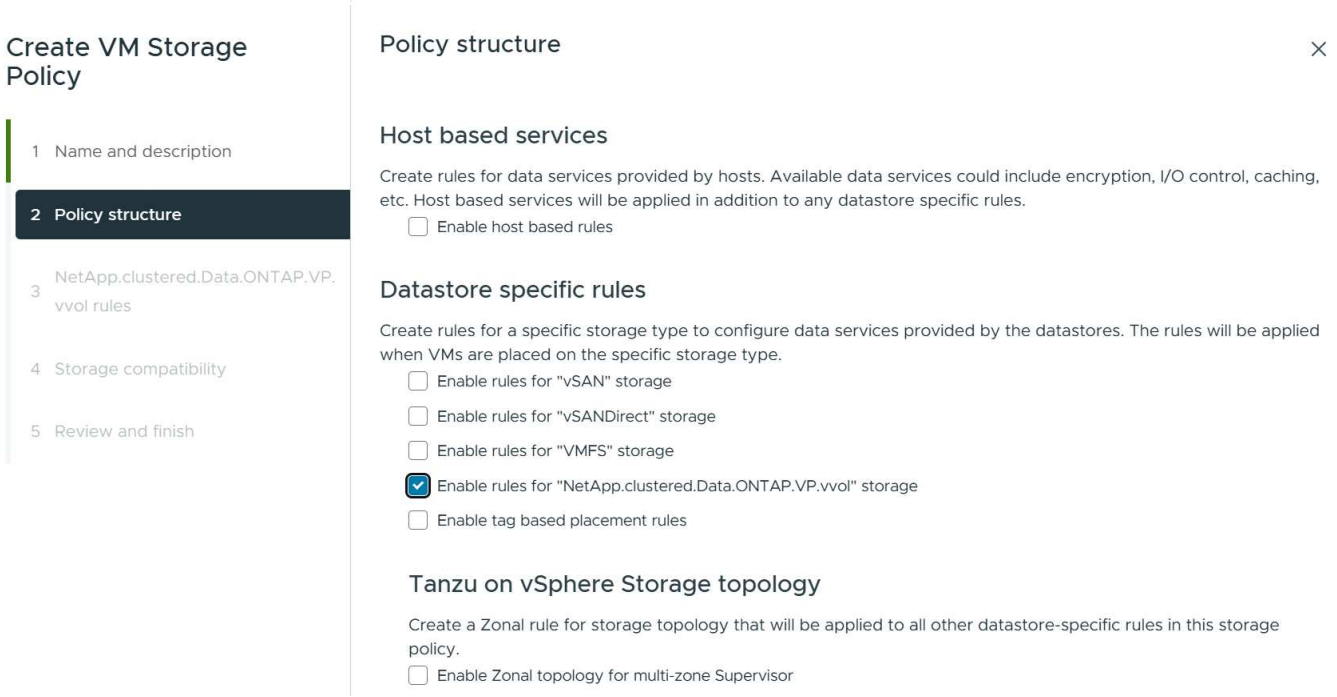
Création de stratégies de stockage de machine virtuelle

Les règles de stockage des machines virtuelles sont utilisées dans vSphere pour gérer les fonctionnalités facultatives telles que le contrôle des E/S du stockage ou le chiffrement vSphere. Ils sont également utilisés avec les vVols pour appliquer des fonctionnalités de stockage spécifiques à la machine virtuelle. Utilisez le type de stockage « NetApp.clustered.Data.ONTAP.VP.vvol ». Voir le lien:vmware-vvols-ontap.html#Best Practices[exemple de configuration réseau avec vVols sur NFS v3] pour un exemple de ceci avec les outils ONTAP VASA Provider. Les règles pour le stockage « NetApp.clustered.Data.ONTAP.VP.VASA10 » doivent être utilisées avec les datastores non basés sur vVols.

Une fois la règle de stockage créée, elle peut être utilisée lors du provisionnement de nouvelles machines virtuelles.


vSphere Client
Search in all environments
Policies and Profiles
VM Storage Policies
VM Customization Specifications
Host Profiles
Compute Policies
Storage Policy Components
VM Storage Policies
CREATE
Quick Filter
Enter value
Name
VC
VM Encryption Policy
vSAN Default Storage Policy
VVol No Requirements Policy
Management Storage Policy - Regular
Management Storage policy - Thin
Management Storage Policy - Large
Management Storage Policy - Stretched
Management Storage Policy - Stretched Lite
Management Storage Policy - Single Node
Management Storage policy - Encryption
Host-local PMem Default Storage Policy
vSAN ESA Default Policy - RAID5
vcf-vc01.ontappmtme.openenglab.netapp.com
Deselect All


Create VM Storage Policy
1 Name and description
2 Policy structure
3 Storage compatibility
4 Review and finish
Name and description
vCenter Server:
VCF-VC01.ONTAPPMTME.OPENENGLAB.NETAPP.COM
Name:
NetApp VM Storage Policy
Description:


Create VM Storage Policy
1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP.vvol rules
4 Storage compatibility
5 Review and finish
Policy structure
Host based services
Create rules for data services provided by hosts. Available data services could include encryption, I/O control, caching, etc. Host based services will be applied in addition to any datastore specific rules.
Enable host based rules
Datastore specific rules
Create rules for a specific storage type to configure data services provided by the datastores. The rules will be applied when VMs are placed on the specific storage type.
Enable rules for "vSAN" storage
Enable rules for "vSANDirect" storage
Enable rules for "VMFS" storage
Enable rules for "NetApp.clustered.Data.ONTAP.VP.vvol" storage
Enable tag based placement rules
Tanzu on vSphere Storage topology
Create a Zonal rule for storage topology that will be applied to all other datastore-specific rules in this storage policy.
Enable Zonal topology for multi-zone Supervisor

Create VM Storage Policy

1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP.vvol rules
4 Storage compatibility
5 Review and finish

NetApp.clustered.Data.ONTAP.VP.vvol rules

Placement | Tags

Platform Type	AFF
Tier	Performance
Space Efficiency	Thin

ADD RULE

QoS IOPS

Create VM Storage Policy

1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP.vvol rules
4 Storage compatibility
5 Review and finish

NetApp.clustered.Data.ONTAP.VP.vvol rules

Placement | Tags

Platform Type	AFF
Tier	Performance
Space Efficiency	Thin
QoS IOPS	REMOVE
MaxThroughput IOPS	10000
MinThroughput IOPS	1000

Create VM Storage Policy

1 Name and description
2 Policy structure
3 NetApp.clustered.Data.ONTAP.VP.vvol rules
4 Storage compatibility
5 Review and finish

Storage compatibility

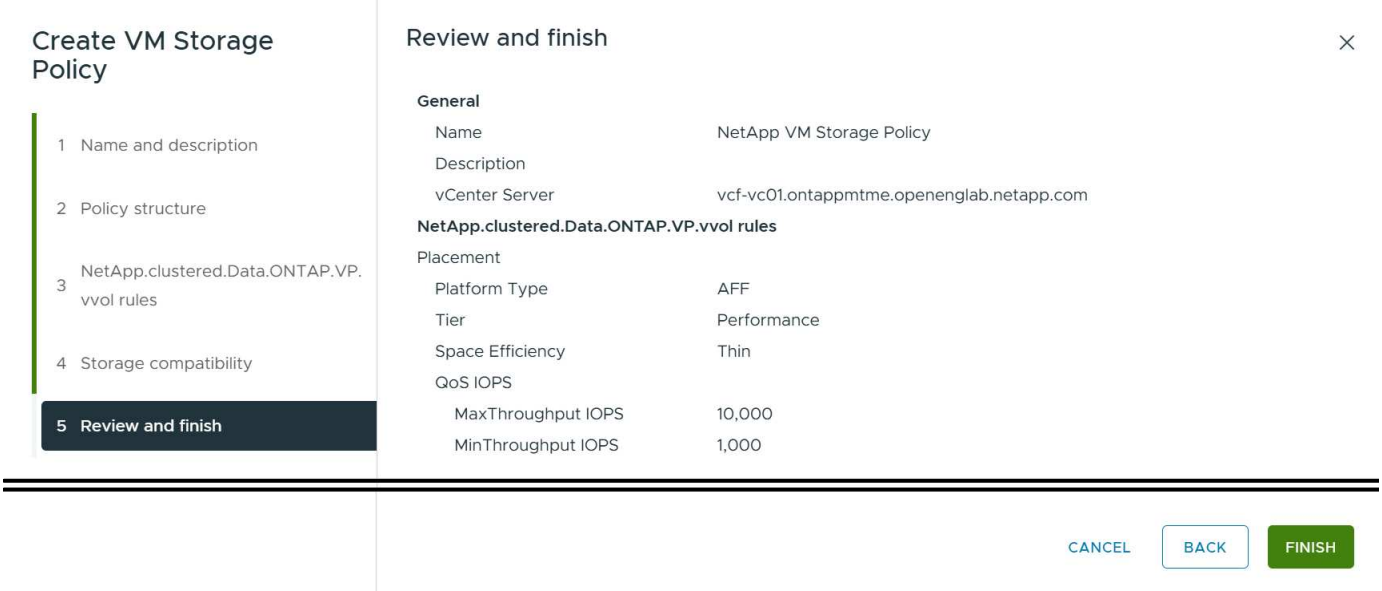
COMPATIBLE | INCOMPATIBLE

Expand datastore clusters

Compatible storage 4 TB (3.8 TB free)

Quick Filter Enter value

Name	Datacenter	Type	Free Space	Capacity	Warnings
NFS_vVols	Raleigh	vVol	3.80 TB	4.00 TB	



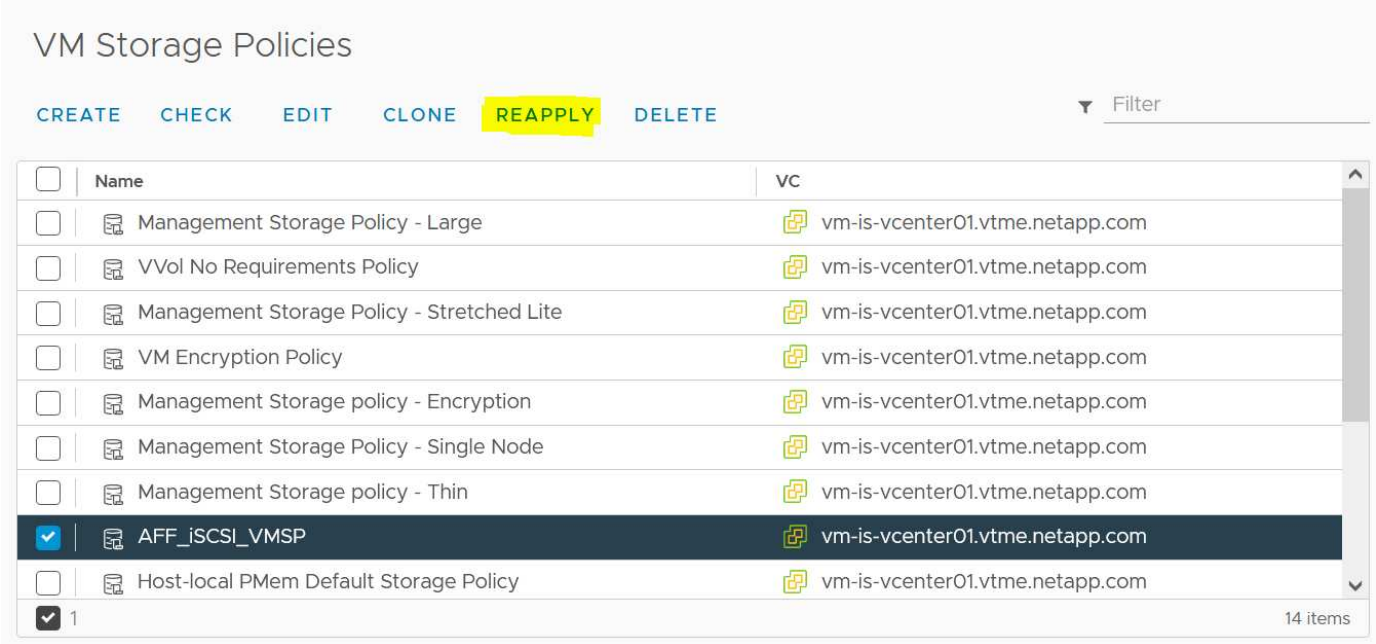
Gestion des performances avec les outils ONTAP

Les outils ONTAP utilisent leur propre algorithme de placement équilibré pour placer un nouveau VVol dans le meilleur FlexVol volume avec des systèmes ASA unifiés ou classiques, ou dans une zone de disponibilité du stockage (SAZ) avec des systèmes ASA r2, dans un datastore vVols. Le placement est basé sur la correspondance entre le stockage de sauvegarde et la règle de stockage des machines virtuelles. Cela permet de s'assurer que le datastore et le stockage de sauvegarde peuvent répondre aux exigences de performances spécifiées.

La modification des capacités de performance telles que les IOPS min et max requiert une certaine attention particulière à la configuration spécifique.

- **Les valeurs min et Max IOPS** peuvent être spécifiées dans une stratégie VM.
  - La modification des IOPS de la règle ne modifiera pas la QoS sur les vVols tant que la règle de machine virtuelle ne sera pas réappliquée aux machines virtuelles qui l'utilisent. Vous pouvez également créer une nouvelle règle avec les IOPS souhaitées et l'appliquer aux machines virtuelles cibles. Il est généralement recommandé de définir simplement des règles de stockage VM distinctes pour les différents niveaux de service et de modifier simplement la règle de stockage VM sur la VM.
  - Les personnalités ASA, ASA r2, AFF et FAS ont des paramètres d'IOPS différents. Les valeurs min et Max sont disponibles sur tous les systèmes Flash. Toutefois, les systèmes non AFF peuvent uniquement utiliser les paramètres Max IOPS.
- Les outils ONTAP créent des règles de QoS individuelles non partagées avec les versions de ONTAP actuellement prises en charge. Par conséquent, chaque VMDK individuel recevra sa propre allocation d'IOPS.

Réapplication de la stratégie de stockage VM



Protection des vVols

Les sections suivantes présentent les procédures et les bonnes pratiques d'utilisation de VMware vVols avec le stockage ONTAP.

Haute disponibilité VASA Provider

Le fournisseur NetApp VASA s'exécute en tant que composant de l'appliance virtuelle, avec le plug-in vCenter et le serveur d'API REST (anciennement Virtual Storage Console [VSC]) et Storage Replication adapter. Si le fournisseur VASA n'est pas disponible, les machines virtuelles utilisant des vVols continueront à s'exécuter. Toutefois, il n'est pas possible de créer de nouveaux datastores vVols et ne peut pas être créé ni lié par vSphere. Cela signifie que les machines virtuelles utilisant des vVols ne peuvent pas être activées car vCenter ne pourra pas demander la création du vVol de swap. De plus, les machines virtuelles en cours d'exécution ne peuvent pas utiliser vMotion pour la migration vers un autre hôte, car les vVols ne peuvent pas être liés au nouvel hôte.

Vasa Provider 7.1 et les versions ultérieures prennent en charge de nouvelles fonctionnalités pour s'assurer que les services sont disponibles dès que nécessaire. Elle comprend de nouveaux processus de surveillance qui surveillent VASA Provider et des services de base de données intégrés. S'il détecte une défaillance, il met à jour les fichiers journaux, puis redémarre automatiquement les services.

L'administrateur vSphere doit configurer une protection supplémentaire en utilisant les mêmes fonctionnalités de disponibilité que celles utilisées pour protéger les autres ordinateurs virtuels stratégiques contre les défaillances logicielles, matérielles hôtes et réseau. Aucune configuration supplémentaire n'est requise sur l'appliance virtuelle pour utiliser ces fonctionnalités ; il vous suffit de les configurer à l'aide des approches vSphere standard. Ils ont été testés et sont pris en charge par NetApp.

VSphere High Availability est facilement configuré pour redémarrer une machine virtuelle sur un autre hôte du cluster hôte en cas de panne. VSphere Fault Tolerance offre une plus grande disponibilité en créant une machine virtuelle secondaire répliquée en continu et capable de prendre le relais à tout moment. Des informations supplémentaires sur ces fonctions sont disponibles dans le "Documentation relative aux outils ONTAP pour VMware vSphere (configuration de la haute disponibilité des outils ONTAP)", Ainsi que la documentation VMware vSphere (recherchez vSphere Availability sous ESXi et vCenter Server).

Le fournisseur VASA des outils ONTAP sauvegarde automatiquement la configuration vVols en temps réel vers des systèmes ONTAP gérés où les informations vVols sont stockées dans les métadonnées de volume FlexVol. Si l'appliance ONTAP Tools devient indisponible, quelle qu'en soit la raison, vous pouvez facilement et rapidement en déployer une nouvelle et importer la configuration. Pour plus d'informations sur les étapes de restauration d'un fournisseur VASA, consultez cet article de la base de connaissances :

"Guide de résolution des incidents VASA Provider"

Réplication vVols

De nombreux clients ONTAP répliquent leurs datastores classiques sur des systèmes de stockage secondaires à l'aide de NetApp SnapMirror, puis utilisent le système secondaire pour restaurer des machines virtuelles individuelles ou la totalité d'un site en cas d'incident. Dans la plupart des cas, les clients utilisent un outil logiciel pour gérer ceci, tel qu'un logiciel de sauvegarde tel que le plug-in NetApp SnapCenter pour VMware vSphere ou une solution de reprise après incident telle que site Recovery Manager de VMware (avec l'adaptateur de réplication du stockage dans les outils ONTAP).

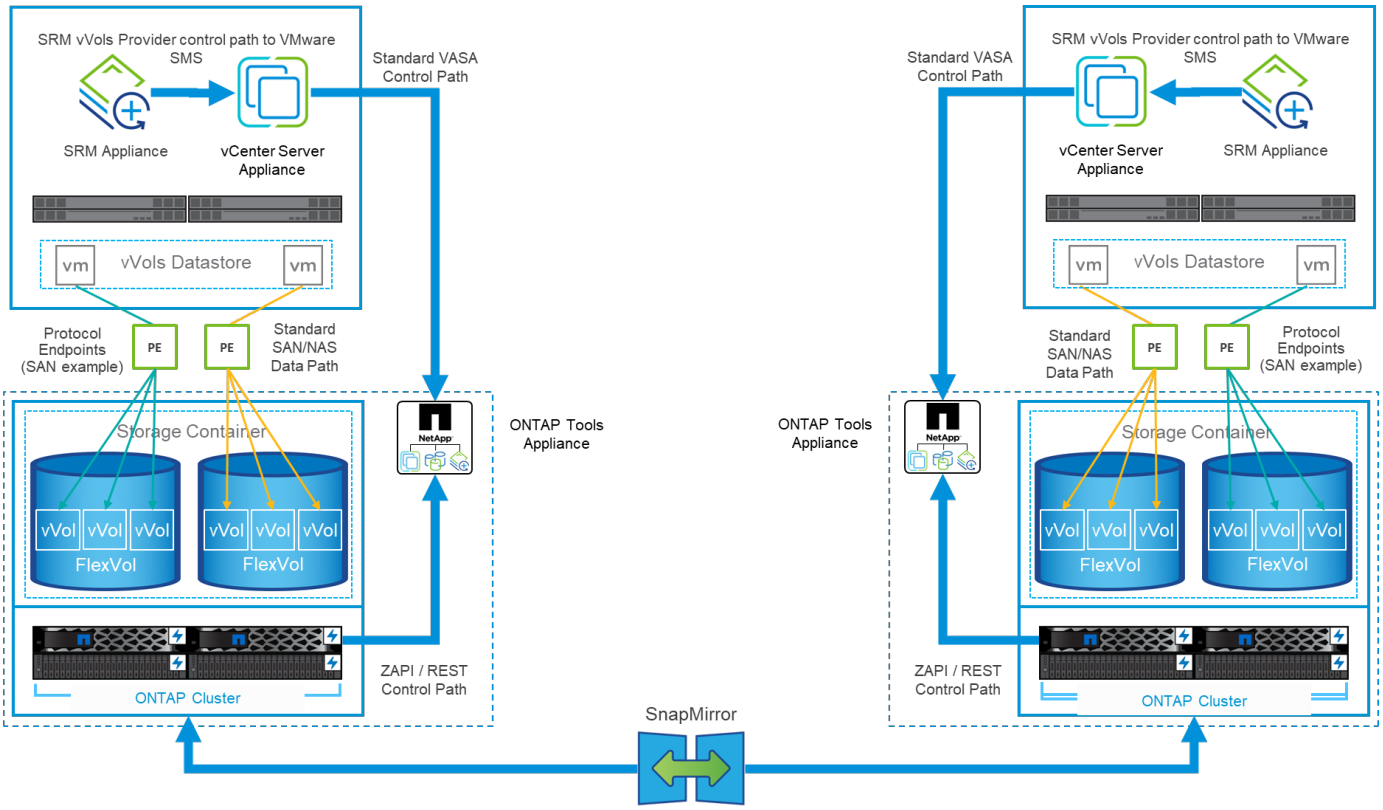
Cette exigence relative à un outil logiciel est encore plus importante pour la gestion de la réplication des vVols. Les fonctionnalités natives permettent de gérer certains aspects (par exemple, les copies Snapshot des vVols gérées par VMware sont déchargées vers ONTAP, qui utilise des clones de fichiers ou de LUN rapides et efficaces). Toutefois, l'orchestration générale est nécessaire pour gérer la réplication et la restauration. Les métadonnées concernant les vVols sont protégées par ONTAP et par le fournisseur VASA, mais des traitements supplémentaires sont nécessaires pour les utiliser sur un site secondaire.

Les outils ONTAP 9.7.1 associés à VMware site Recovery Manager (SRM) 8.3 ont également pris en charge la reprise après incident et l'orchestration des flux de travail de migration en tirant parti de la technologie NetApp SnapMirror.

Dans la version initiale de la prise en charge de SRM avec les outils ONTAP 9.7.1, il était nécessaire de pré-créer des volumes FlexVol et d'activer la protection SnapMirror avant de les utiliser comme volumes de sauvegarde d'un datastore vVols. À partir des outils ONTAP 9.10, ce processus n'est plus nécessaire. Vous pouvez désormais ajouter la protection SnapMirror aux volumes de sauvegarde existants et mettre à jour les règles de stockage de vos machines virtuelles afin de bénéficier d'une gestion basée sur des règles avec reprise après incident, orchestration de la migration et automatisation intégrées à SRM.

Actuellement, VMware SRM est la seule solution d'automatisation de la migration et de la reprise après incident pour les vVols pris en charge par NetApp. Les outils ONTAP vérifient l'existence d'un serveur SRM 8.3 ou version ultérieure enregistré dans votre vCenter avant de vous permettre d'activer la réplication vVols, Vous pouvez exploiter les API REST d'outils ONTAP pour créer vos propres services.

Réplication de vVols avec SRM



Support MetroCluster

Bien que les outils ONTAP ne soient pas capables de déclencher un basculement MetroCluster, ils prennent en charge les systèmes NetApp MetroCluster pour les vVols soutenant les volumes dans une configuration vMSC (vSphere Metro Storage Cluster) uniforme. Le basculement d'un système MetroCluster est géré de la manière habituelle.

Même si NetApp SnapMirror Business Continuity (SM-BC) peut également servir de base pour une configuration vMSC, il n'est pas pris en charge avec les vVols.

Pour plus d'informations sur NetApp MetroCluster, consultez ces guides :

"*TR-4689 Architecture et conception de la solution MetroCluster IP*"

"*TR-4705 Architecture et conception de la solution NetApp MetroCluster*"

"*VMware KB 2031038 prise en charge de VMware vSphere avec NetApp MetroCluster*"

Présentation de la sauvegarde vVols

Il existe plusieurs approches pour protéger les machines virtuelles, telles que l'utilisation d'agents de sauvegarde invités, la connexion de fichiers de données VM à un proxy de sauvegarde ou l'utilisation d'API définies telles que VMware VADP. Les vVols peuvent être protégées à l'aide des mêmes mécanismes et de nombreux partenaires NetApp prennent en charge les sauvegardes de machines virtuelles, y compris les vVols.

Comme mentionné précédemment, les snapshots gérés par VMware vCenter sont déchargés dans des clones de fichiers/LUN ONTAP rapides et compacts. Elles peuvent être utilisées pour des sauvegardes rapides et manuelles, mais vCenter limite le nombre de snapshots à 32. Vous pouvez utiliser vCenter pour créer des snapshots et restaurer les données selon vos besoins.

À partir du plug-in SnapCenter pour VMware vSphere (SCV) 4.6 utilisé conjointement avec les outils ONTAP 9.10 et versions ultérieures, ajoute la prise en charge de la sauvegarde et de la restauration cohérentes après panne des machines virtuelles basées sur vVols exploitant les snapshots de volume ONTAP FlexVol avec prise en charge de la réplication SnapMirror et SnapVault. Jusqu'à 1023 copies Snapshot sont prises en charge par volume. SCV peut également stocker davantage de copies Snapshot avec une conservation plus longue sur des volumes secondaires à l'aide de SnapMirror avec une règle de copie miroir.

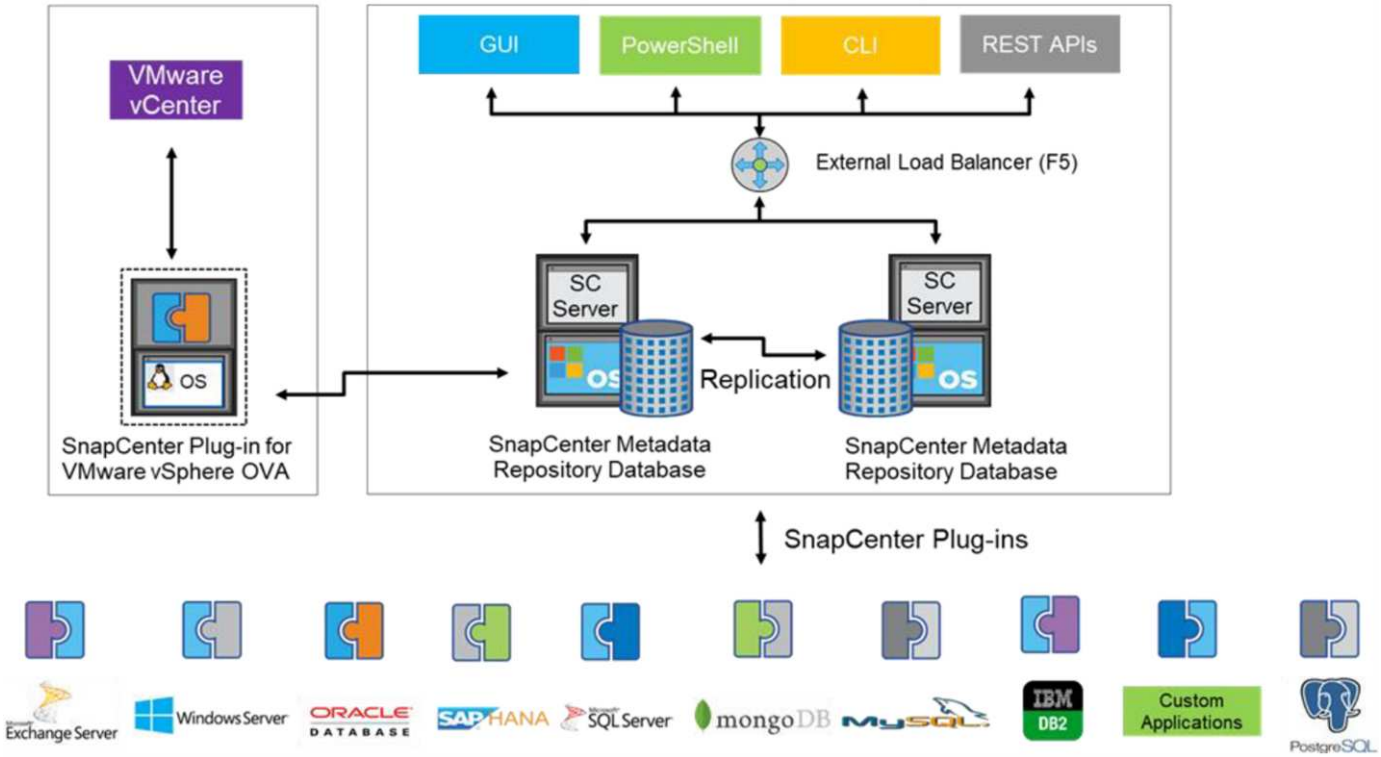
La prise en charge de vSphere 8.0 a été introduite avec SCV 4.7, qui utilisait une architecture de plug-ins locaux isolée. La prise en charge de vSphere 8.0U1 a été ajoutée à SCV 4.8, qui a entièrement migré vers la nouvelle architecture de plug-ins distants.

VVols Backup avec le plug-in SnapCenter pour VMware vSphere

Avec NetApp SnapCenter, vous pouvez désormais créer des groupes de ressources pour les vVols à partir de balises et/ou de dossiers afin de tirer automatiquement parti des snapshots FlexVol d'ONTAP pour les machines virtuelles basées sur vVols. Cela vous permet de définir des services de sauvegarde et de restauration qui protègent automatiquement les machines virtuelles lorsqu'elles sont provisionnées dynamiquement au sein de votre environnement.

Le plug-in SnapCenter pour VMware vSphere est déployé en tant qu'appliance autonome enregistrée en tant qu'extension vCenter, gérée via l'interface utilisateur vCenter ou via les API REST pour l'automatisation des services de sauvegarde et de restauration.

Architecture SnapCenter



Comme les autres plug-ins SnapCenter ne prennent pas encore en charge les vVols au moment de la rédaction de ce document, nous nous concentrerons sur le modèle de déploiement autonome présenté dans ce document.

Étant donné que SnapCenter utilise les copies Snapshot ONTAP FlexVol, il n'y a pas de surcharge placée sur vSphere, ni de réduction des performances comme on peut le voir avec les machines virtuelles traditionnelles

utilisant les snapshots gérés par vCenter. De plus, comme la fonctionnalité de SCV est exposée via les API REST, il est facile de créer des workflows automatisés à l'aide d'outils tels que VMware Aria Automation, Ansible, Terraform et pratiquement tous les autres outils d'automatisation capables d'utiliser des API REST standard.

Pour plus d'informations sur les API REST de SnapCenter, reportez-vous à la section "Présentation des API REST"

Pour plus d'informations sur le plug-in SnapCenter pour les API REST VMware vSphere, consultez la section "Plug-in SnapCenter pour les API REST VMware vSphere"

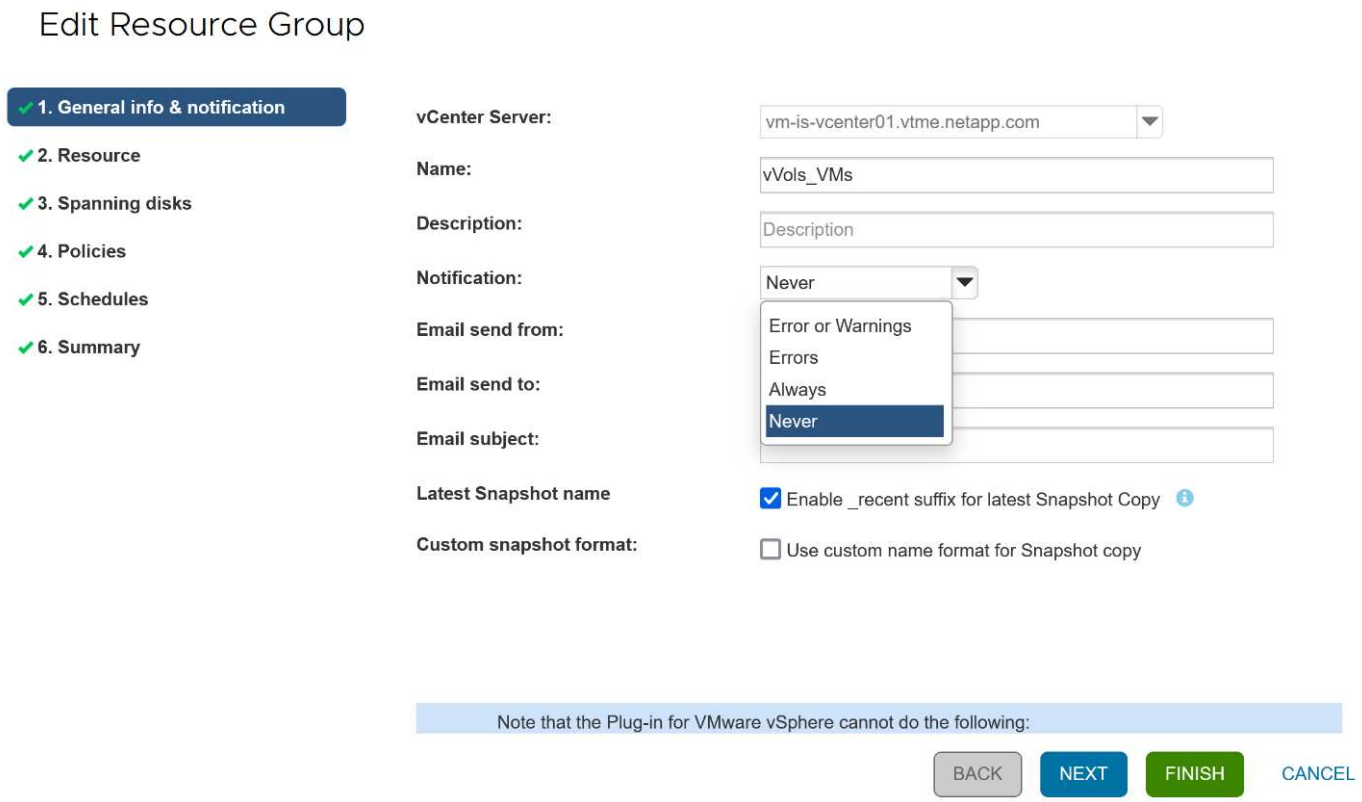
Et des meilleures pratiques

Les bonnes pratiques suivantes peuvent vous aider à tirer le meilleur parti de votre déploiement SnapCenter.

• SCV prend en charge les rôles RBAC vCenter Server et ONTAP RBAC et inclut des rôles vCenter prédéfinis qui sont automatiquement créés pour vous lorsque le plug-in est enregistré. Vous pouvez en savoir plus sur les types de RBAC pris en charge "ici."
  ◦ Utilisez l'interface utilisateur de vCenter pour attribuer l'accès au compte le moins privilégié à l'aide des rôles prédéfinis décrits "ici".
  ◦ Si vous utilisez SCV avec le serveur SnapCenter, vous devez attribuer le rôle *SnapCenter_Admin*.
  ◦ ONTAP RBAC fait référence au compte utilisateur utilisé pour ajouter et gérer les systèmes de stockage utilisés par SCV. ONTAP RBAC ne s'applique pas aux sauvegardes basées sur vVols. En savoir plus sur ONTAP RBAC et SCV "ici".

• Répliquez vos jeux de données de sauvegarde sur un second système à l'aide de SnapMirror pour créer des répliques complètes des volumes source. Comme mentionné précédemment, vous pouvez également utiliser des règles de copie miroir pour la conservation à long terme des données de sauvegarde, indépendamment des paramètres de conservation des snapshots du volume source. Les deux mécanismes sont pris en charge avec vVols.
• Étant donné que SCV requiert également les outils ONTAP pour la fonctionnalité VMware vSphere for vVols, vérifiez toujours la compatibilité des versions avec l'outil IMT (Interoperability Matrix Tool) de NetApp
• Si vous utilisez la réplication vVols avec VMware SRM, tenez compte de vos objectifs RPO et de votre planification de sauvegarde
• Concevez vos règles de sauvegarde avec des paramètres de conservation qui répondent aux objectifs de point de restauration (RPO) définis par votre entreprise.
• Configurez les paramètres de notification de vos groupes de ressources pour qu'ils soient informés de l'état lors de l'exécution des sauvegardes (voir la figure 10 ci-dessous).

Options de notification de groupe de ressources



Commencer à utiliser SCV à l'aide de ces documents

"En savoir plus sur le plug-in SnapCenter pour VMware vSphere"

"Déployez le plug-in SnapCenter pour VMware vSphere"

Dépannage

Plusieurs ressources de dépannage sont disponibles avec des informations supplémentaires.

Site de support NetApp

Outre plusieurs articles de la base de connaissances sur les produits de virtualisation NetApp, le site de support NetApp offre également une page d'accueil pratique pour "Les outils ONTAP pour VMware vSphere" le produit. Ce portail propose des liens vers des articles, des téléchargements, des rapports techniques et des discussions sur les solutions VMware sur la communauté NetApp. Il est disponible à l'adresse suivante :

"*Site de support NetApp*"

Vous trouverez une documentation supplémentaire sur les solutions ici :

"*Solutions NetApp de virtualisation avec VMware de Broadcom*"

Dépannage du produit

Les différents composants des outils ONTAP, tels que le plug-in vCenter, VASA Provider et Storage Replication adapter sont tous documentés dans le référentiel de documents NetApp. Cependant, chacun d'entre eux dispose d'une sous-section distincte de la base de connaissances et peut avoir des procédures de dépannage

spécifiques. Ils répondent aux problèmes les plus courants rencontrés avec le fournisseur VASA.

Problèmes liés à l'interface utilisateur de VASA Provider

Il arrive que le client Web vCenter vSphere rencontre des problèmes avec les composants Serenity, ce qui empêche l'affichage des éléments de menu VASA Provider for ONTAP. Consultez la section résolution des problèmes d'enregistrement de VASA Provider dans le Guide de déploiement ou cette base de connaissances "article".

Échec du provisionnement du datastore vVols

Il arrive parfois que les services vCenter prennent du temps lors de la création du datastore vVols. Pour le corriger, redémarrez le service vmware-sps et remontez le datastore vVols à l'aide des menus vCenter (stockage > Nouveau datastore). Ceci est couvert par les échecs de provisionnement du datastore vVols avec vCenter Server 6.5 dans le Guide d'administration.

La mise à niveau d'Unified Appliance ne parvient pas à monter l'ISO

En raison d'un bogue dans vCenter, le montage de l'ISO utilisé pour mettre à niveau l'appliance unifiée d'une version à l'autre peut échouer. Si l'ISO peut être attaché à l'appliance dans vCenter, suivez la procédure de cette base de connaissances "article" à résoudre.

VMware site Recovery Manager et ONTAP

Restauration de site en direct VMware avec ONTAP

ONTAP est une solution de stockage de premier plan pour VMware vSphere et, plus récemment, Cloud Foundation, depuis l'introduction d'ESX dans les centres de données modernes il y a plus de deux décennies. NetApp continue d'introduire des systèmes innovants, tels que la dernière génération de la série ASA A, ainsi que des fonctionnalités telles que la synchronisation active SnapMirror . Ces avancées simplifient la gestion, améliorent la résilience et réduisent le coût total de possession (TCO) de votre infrastructure informatique.

Ce document présente la solution ONTAP pour VMware Live Site Recovery (VLSR), anciennement connu sous le nom de Site Recovery Manager (SRM), le logiciel de reprise après sinistre (DR) leader du secteur de VMware, y compris les dernières informations sur les produits et les meilleures pratiques pour rationaliser le déploiement, réduire les risques et simplifier la gestion continue.

> Cette documentation remplace le rapport technique précédemment publié *TR-4900 : VMware Site Recovery Manager avec ONTAP*

Les meilleures pratiques complètent d'autres documents, tels que des guides et des outils de compatibilité. Ils sont développés en fonction de tests effectués en laboratoire et d'une vaste expérience sur le terrain par les ingénieurs et les clients NetApp. Dans certains cas, les meilleures pratiques recommandées peuvent ne pas être adaptées à votre environnement. Cependant, ce sont généralement les solutions les plus simples qui répondent aux besoins des plus clients.

Ce document porte sur les fonctionnalités des dernières versions de ONTAP 9 utilisées conjointement avec les outils ONTAP pour VMware vSphere 10.4 (notamment NetApp Storage Replication adapter [SRA] et VASA Provider [VP]), ainsi que sur VMware Live site Recovery 9.

Pourquoi utiliser ONTAP avec VLSR ou SRM ?

Les plates-formes de gestion de données NetApp optimisées par ONTAP font partie des solutions de stockage les plus largement adoptées pour VLSR. Les raisons sont nombreuses : une plate-forme de gestion de données sécurisée, hautes performances et unifiée (NAS et SAN ensemble) qui offre une efficacité de stockage définissant le secteur, une multilocation, des contrôles de qualité de service, une protection des données avec des instantanés économes en espace et une réplication avec SnapMirror. Le tout s'appuyant sur l'intégration multicloud hybride native pour la protection des charges de travail VMware et une pléthore d'outils d'automatisation et d'orchestration à portée de main.

Lorsque vous utilisez SnapMirror pour la réplication basée sur une baie, vous bénéficiez de l'une des technologies les plus éprouvées et les plus matures d'ONTAP. SnapMirror vous offre l'avantage de transferts de données sécurisés et très efficaces, en copiant uniquement les blocs de système de fichiers modifiés, et non des machines virtuelles ou des banques de données entières. Même ces blocs bénéficient d'économies d'espace, telles que la déduplication, la compression et le compactage. Les systèmes ONTAP modernes utilisent désormais SnapMirror indépendant de la version, ce qui vous permet de sélectionner vos clusters source et de destination avec flexibilité. SnapMirror est véritablement devenu l'un des outils les plus puissants disponibles pour la reprise après sinistre.

Que vous utilisiez des magasins de données traditionnels NFS, iSCSI ou Fibre Channel (désormais avec prise en charge des magasins de données vVols), VLSR fournit une offre propriétaire robuste qui exploite le meilleur des fonctionnalités ONTAP pour la reprise après sinistre ou la planification et l'orchestration de la migration du centre de données.

Comment VLSR exploite ONTAP 9

VLSR exploite les technologies avancées de gestion des données des systèmes ONTAP en l'intégrant aux outils ONTAP pour VMware vSphere, une appliance virtuelle qui englobe trois composants principaux :

- Le plug-in vCenter des outils ONTAP, précédemment appelé Virtual Storage Console (VSC), simplifie les fonctionnalités d'efficacité et de gestion du stockage, améliore la disponibilité et réduit les coûts de stockage et les coûts d'exploitation, que vous utilisiez SAN ou NAS. Il s'appuie sur les bonnes pratiques pour le provisionnement des datastores et optimise les paramètres d'hôte ESXi pour les environnements de stockage NFS et bloc. Pour tous ces avantages, NetApp recommande ce plug-in lors de l'utilisation de vSphere avec les systèmes exécutant ONTAP.
- Les outils ONTAP VASA Provider prennent en charge le framework VMware vStorage APIs for Storage Awareness (VASA). Vasa Provider connecte vCenter Server avec ONTAP pour faciliter le provisionnement et la surveillance du stockage des machines virtuelles. Cela a permis la prise en charge des volumes virtuels VMware (vVols) et la gestion des politiques de stockage des machines virtuelles et des performances des vVols des machines virtuelles individuelles. Il fournit également des alarmes pour la surveillance de la capacité et la conformité avec les profils.
- SRA est utilisée en association avec VLSR pour gérer la réplication des données des machines virtuelles entre les sites de production et de reprise après incident pour les datastores VMFS et NFS traditionnels, et pour les tests non disruptives des répliques de DR. Il permet d'automatiser les tâches de détection, de restauration et de reprotection. Il comprend à la fois un serveur SRA et des adaptateurs SRA pour le serveur Windows SRM et le dispositif VLSR.

Après avoir installé et configuré les adaptateurs SRA sur le serveur VLSR pour protéger les banques de données non vVols, vous pouvez commencer la tâche de configuration de votre environnement vSphere pour la reprise après sinistre.

SRA est une interface de commande et de contrôle pour le serveur VLSR afin de gérer les volumes ONTAP FlexVol qui contiennent vos machines virtuelles VMware, ainsi que la réplication SnapMirror qui les protège.

VLSR peut tester votre plan DR sans interruption à l'aide de la technologie propriétaire FlexClone de NetApp pour créer des clones quasi instantanés de vos banques de données protégées sur votre site DR. VLSR crée un bac à sable pour effectuer des tests en toute sécurité afin que votre organisation et vos clients soient protégés en cas de véritable sinistre, vous donnant confiance dans la capacité de votre organisation à exécuter un basculement en cas de sinistre.

En cas d'incident véritable ou même de migration planifiée, VLSR vous permet d'envoyer les modifications de dernière minute au jeu de données via une mise à jour SnapMirror finale (si vous le souhaitez). Il interrompt ensuite le miroir et monte le datastore sur vos hôtes de reprise après incident. À ce stade, vos machines virtuelles peuvent être automatiquement alimentées dans l'ordre de votre stratégie prédéfinie.

> Bien que les systèmes ONTAP vous permettent de coupler des SVM au sein du même cluster pour la réplication SnapMirror, ce scénario n'est pas testé et certifié avec VLSR. Par conséquent, il est recommandé d'utiliser uniquement des SVM provenant de différents clusters lors de l'utilisation de VLSR.

VLSR avec ONTAP et autres cas d'utilisation : cloud hybride et migration

L'intégration de votre déploiement VLSR aux fonctionnalités avancées de gestion des données ONTAP permet une évolutivité et des performances considérablement améliorées par rapport aux options de stockage local. Mais plus que cela, il apporte la flexibilité du cloud hybride. Le cloud hybride vous permet d'économiser de l'argent en hiérarchisant les blocs de données inutilisés de votre baie hautes performances vers votre hyperscaler préféré à l'aide de FabricPool, qui peut être un magasin S3 sur site tel que NetApp StorageGRID. Vous pouvez également utiliser SnapMirror pour les systèmes Edge avec ONTAP Select défini par logiciel ou DR basé sur le cloud à l'aide de Cloud Volumes ONTAP (CVO) ou "Stockage NetApp sur Equinix Metal®" , ou d'autres services ONTAP hébergés.

Vous pouvez ensuite effectuer un basculement de test dans le data Center d'un fournisseur de services clouds avec une empreinte de stockage proche de zéro grâce à FlexClone. La protection de votre entreprise peut à présent être plus économique que jamais.

VLSR peut également être utilisé pour exécuter des migrations planifiées en utilisant SnapMirror pour transférer efficacement vos machines virtuelles d'un data Center à un autre ou même au sein d'un même data Center, que vous le soyez propriétaire ou via plusieurs fournisseurs de services partenaires NetApp.

Bonnes pratiques de déploiement

Les sections suivantes présentent les meilleures pratiques de déploiement avec ONTAP et VMware SRM.

Utiliser la dernière version des outils ONTAP 10

Les outils ONTAP 10 offrent des améliorations significatives par rapport aux versions précédentes, notamment :

- basculement de test 8 fois plus rapide*
- nettoyage et reprotection 2 fois plus rapides*
- basculement 32 % plus rapide*
- Évolutivité accrue
- Prise en charge native des mises en page de sites partagés

*Ces améliorations sont basées sur des tests internes et peuvent varier en fonction de votre environnement.

Disposition des SVM et segmentation pour la colocation sécurisée

Avec ONTAP, le concept de machine virtuelle de stockage (SVM) offre une segmentation stricte dans les environnements mutualisés sécurisés. Les utilisateurs des SVM situés sur un SVM ne peuvent ni accéder aux ressources d'un autre ni les gérer. De cette façon, vous pouvez exploiter la technologie ONTAP en créant des SVM distincts pour différentes unités commerciales qui gèrent leurs propres flux de travail SRM sur le même cluster, pour une efficacité globale supérieure du stockage.

Envisagez de gérer ONTAP avec des comptes SVM-scoped et des LIF de management SVM pour non seulement améliorer les contrôles de sécurité, mais aussi améliorer les performances. Les performances sont supérieures par nature lorsque des connexions SVM-scoped sont utilisées, car SRA n'est pas nécessaire pour traiter toutes les ressources d'un cluster entier, y compris les ressources physiques. Il ne doit plutôt comprendre que les ressources logiques qui sont extraites vers la SVM particulière.

Meilleures pratiques pour la gestion des systèmes ONTAP 9

Comme mentionné précédemment, il est possible de gérer des clusters ONTAP avec des identifiants cluster ou SVM évalués et des LIF de gestion. Pour des performances optimales, il peut être intéressant d'utiliser des identifiants SVM- scoped lorsque vous n'utilisez pas les vVols. Cependant, ce faisant, vous devriez être conscient de certaines exigences, et que vous perdez certaines fonctionnalités.

- Le compte SVM vsadmin par défaut ne dispose pas du niveau d'accès requis pour effectuer les tâches des outils ONTAP Il faut donc créer un compte SVM. "Configurer les rôles et privilèges des utilisateurs ONTAP" Utilisation du fichier JSON inclus. Il peut être utilisé pour les comptes évalués au niveau du SVM ou du cluster.
- Comme le plug-in de l'interface utilisateur vCenter, le fournisseur VASA et le serveur SRA sont tous des microservices entièrement intégrés, vous devez ajouter du stockage à l'adaptateur SRA dans SRM de la même manière que vous ajoutez du stockage dans l'interface utilisateur vCenter pour les outils ONTAP. Sinon, le serveur SRA pourrait ne pas reconnaître les requêtes envoyées depuis SRM via l'adaptateur SRA.
- La vérification du chemin NFS n'est pas effectuée en cas d'utilisation d'identifiants SVM-scoped, sauf si vous êtes le premier dans ONTAP Tools Manager et si vous "clusters intégrés" les associez à vCenters. Car l'emplacement physique est logiquement extrait du SVM. Cela ne pose pas de problème, car les systèmes ONTAP modernes ne subissent plus de déclin perceptible des performances lors de l'utilisation de chemins indirects.
- Il est possible que les économies d'espace réalisées grâce à l'efficacité du stockage ne soient pas signalées.
- Lorsqu'ils sont pris en charge, les miroirs de partage de charge ne peuvent pas être mis à jour.
- Il est possible que la connexion EMS ne soit pas effectuée sur des systèmes ONTAP gérés avec des identifiants évalués par SVM.

Meilleures pratiques opérationnelles

Les sections suivantes présentent les meilleures pratiques opérationnelles pour VMware SRM et le stockage ONTAP.

Datastores et protocoles

- Si possible, utilisez toujours les outils ONTAP pour provisionner les datastores et les volumes. Cela vérifie que les volumes, les chemins de jonction, les LUN, les igroups, les règles d'exportation, et d'autres paramètres sont configurés de manière compatible.

- SRM prend en charge iSCSI, Fibre Channel et NFS version 3 avec ONTAP 9 lors de l'utilisation d'une réplication basée sur les baies via SRA. SRM ne prend pas en charge la réplication basée sur la baie pour NFS version 4.1 avec des datastores traditionnels ou vvols.
- Pour confirmer la connectivité, vérifiez toujours que vous pouvez monter et démonter un nouveau datastore test sur le site de reprise sur incident à partir du cluster ONTAP de destination. Testez chaque protocole que vous envisagez d'utiliser pour la connectivité du datastore. L'une des meilleures pratiques est d'utiliser les outils ONTAP pour créer votre datastore de test, car elle effectue toutes les automatisations du datastore telles que dirigées par SRM.
- Les protocoles SAN doivent être homogènes pour chaque site. Vous pouvez combiner les protocoles NFS et SAN, mais les protocoles SAN ne doivent pas être combinés dans un même site. Par exemple, vous pouvez utiliser FCP sur Le site A et iSCSI sur le site B. vous ne devez pas utiliser FCP ou iSCSI sur le site A.
- Les guides précédents ont recommandé de créer la LIF pour la localisation des données. C'est-à-dire toujours monter un datastore à l'aide d'une LIF située sur le nœud qui détient physiquement le volume. Bien que ce soit toujours la meilleure pratique, ce n'est plus une exigence dans les versions modernes de ONTAP 9. Dans la mesure du possible, et si des informations d'identification avec périmètre du cluster sont fournies, les outils ONTAP choisissent toujours d'équilibrer la charge entre les LIF locales aux données, mais il ne s'agit pas d'une exigence de haute disponibilité ou de performance.
- ONTAP 9 peut être configuré pour supprimer automatiquement les snapshots afin de préserver la disponibilité en cas de manque d'espace lorsque la taille automatique ne peut pas fournir une capacité d'urgence suffisante. Le paramètre par défaut de cette fonctionnalité ne supprime pas automatiquement les snapshots créés par SnapMirror. Si des snapshots SnapMirror sont supprimés, NetApp SRA ne peut pas inverser et resynchroniser la réplication pour le volume affecté. Pour empêcher ONTAP de supprimer des snapshots SnapMirror, configurez la fonction de suppression automatique des snapshots sur « essayer ».

```
snap autodelete modify -volume -commitment try
```

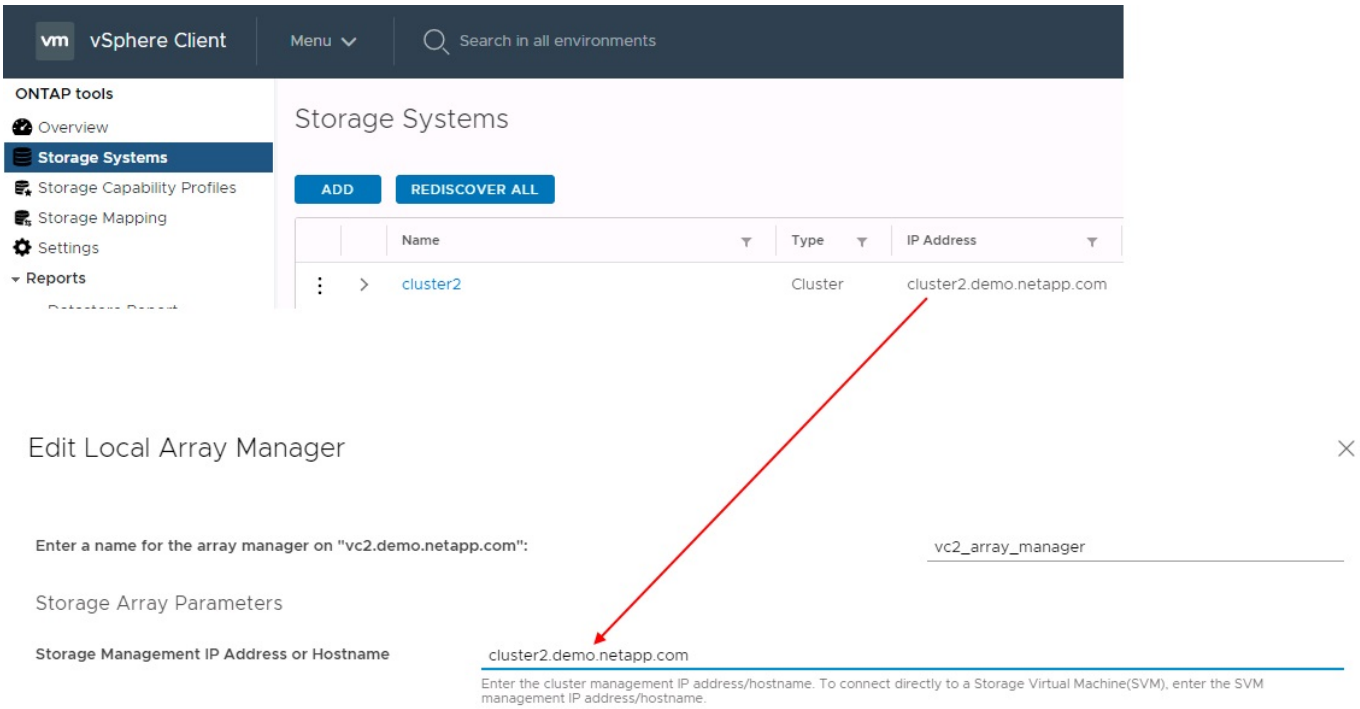
- La taille automatique du volume doit être définie sur `grow` pour les volumes contenant des datastores SAN et `grow_shrink` pour les datastores NFS. Pour en savoir plus sur ce sujet"Configurez les volumes pour qu'ils augmentent ou réduisent automatiquement leur taille", rendez-vous sur .
- SRM fonctionne mieux lorsque le nombre de datastores et donc les groupes de protection sont limités dans vos plans de reprise d'activité. Par conséquent, vous devez envisager d'optimiser la densité des machines virtuelles dans les environnements protégés par SRM où le RTO est essentiel.
- Utilisez Distributed Resource Scheduler (DRS) pour équilibrer la charge sur vos clusters ESXi protégés et de récupération. N'oubliez pas que si vous prévoyez de revenir en arrière, lorsque vous exécutez une reprotection, les clusters précédemment protégés deviennent les nouveaux clusters de récupération. Le DRS contribue à équilibrer le placement dans les deux sens.
- Dans la mesure du possible, évitez d'utiliser la personnalisation IP avec SRM car cela peut augmenter votre RTO.

À propos des paires de baies

Un gestionnaire de matrices est créé pour chaque paire de matrices. Avec les outils SRM et ONTAP, chaque association de baie s'effectue au sein d'un SVM, même si vous utilisez les identifiants du cluster. Vous pouvez ainsi segmenter les flux de travail de reprise après incident entre des locataires, en fonction des SVM qu'ils ont affectés à la gestion. Vous pouvez créer plusieurs gestionnaires de baies pour un cluster donné, qui peuvent être asymétriques. Vous pouvez « Fan-Out » ou « Fan-In » sur différents clusters ONTAP 9. Par exemple, il peut y avoir des SVM-A et SVM-B dans le Cluster-1 en cours de réplication vers SVM-C dans le Cluster-2,

SVM-D dans le Cluster-3 ou vice-versa.

Lors de la configuration des paires de baies dans SRM, vous devez toujours les ajouter à SRM de la même manière que vous les avez ajoutés à ONTAP Tools : autrement dit, ils doivent utiliser le même nom d'utilisateur, mot de passe et LIF de gestion. Cette exigence garantit que SRA communique correctement avec la baie. La copie d'écran suivante montre comment un cluster peut s'afficher dans les outils ONTAP et comment il peut être ajouté à un gestionnaire de baies.



À propos des groupes de réplication

Les groupes de réplication contiennent des ensembles logiques de machines virtuelles qui sont restaurées ensemble. Étant donné que la réplication SnapMirror de ONTAP se produit au niveau du volume, toutes les machines virtuelles d'un volume se trouvent dans le même groupe de réplication.

Il existe plusieurs facteurs à prendre en compte dans les groupes de réplication et dans la manière dont vous distribuez les machines virtuelles sur les volumes FlexVol. Le regroupement de machines virtuelles similaires dans un même volume peut améliorer l'efficacité du stockage avec les systèmes ONTAP plus anciens qui n'offrent pas de déduplication au niveau de l'agrégat. Cependant, ce regroupement augmente la taille du volume et réduit la simultanéité E/S du volume. Les systèmes ONTAP modernes offrent un équilibre parfait entre performance et efficacité du stockage en distribuant les machines virtuelles entre les volumes FlexVol au sein d'un même agrégat. La déduplication au niveau de l'agrégat améliore la parallélisation des E/S sur plusieurs volumes. Vous pouvez restaurer des VM dans les volumes simultanément, car un groupe de protection (voir ci-dessous) peut contenir plusieurs groupes de réplication. L'inconvénient de cette disposition est que les blocs peuvent être transmis plusieurs fois via le réseau, car SnapMirror ne prend pas en compte la déduplication dans l'agrégat.

Dernier point à prendre en compte pour les groupes de réplication : chacun d'entre eux est, par nature, un groupe de cohérence logique (à ne pas confondre avec les groupes de cohérence SRM). En effet, toutes les machines virtuelles du volume sont transférées ensemble à l'aide du même snapshot. Ainsi, si vous disposez de machines virtuelles qui doivent être cohérentes les unes avec les autres, envisagez de les stocker dans le même FlexVol.

À propos des groupes de protection

Les groupes de protection définissent les VM et les datastores dans des groupes restaurés à partir du site protégé. Le site protégé est là où existent les VM configurées dans un groupe de protection pendant les opérations stables. Il est important de noter que même si SRM peut afficher plusieurs gestionnaires de baies pour un groupe de protection, un groupe de protection ne peut pas s'étendre sur plusieurs gestionnaires de baies. Pour cette raison, vous ne devez pas couvrir les fichiers de machine virtuelle sur plusieurs datastores sur différents SVM.

À propos des plans de reprise

Les plans de reprise définissent les groupes de protection qui sont restaurés au cours du même processus. Plusieurs groupes de protection peuvent être configurés dans le même plan de reprise. Par ailleurs, pour activer davantage d'options pour l'exécution des plans de reprise, un seul groupe de protection peut être inclus dans plusieurs plans de restauration.

Les plans de restauration permettent aux administrateurs SRM de définir les flux de travail de restauration en affectant des VM à un groupe de priorité compris entre 1 (le plus élevé) et 5 (le plus faible), dont la valeur par défaut est 3 (moyen). Au sein d'un groupe de priorités, les VM peuvent être configurés pour les dépendances.

Par exemple, votre entreprise peut disposer d'une application stratégique de niveau 1 qui repose sur un serveur Microsoft SQL pour sa base de données. Vous décidez donc de placer vos machines virtuelles dans le groupe de priorité 1. Au sein du groupe de priorité 1, vous commencez à planifier la commande afin d'obtenir des services. Vous souhaitez probablement que votre contrôleur de domaine Microsoft Windows démarre avant votre serveur Microsoft SQL, qui doit être en ligne avant votre serveur d'applications, et ainsi de suite. Vous devez ajouter toutes ces machines virtuelles au groupe de priorité, puis définir les dépendances, car elles ne s'appliquent qu'à un groupe de priorité donné.

NetApp recommande fortement de travailler avec vos équipes en charge des applications pour comprendre l'ordre des opérations requises dans un scénario de basculement et pour élaborer vos plans de reprise en conséquence.

Tester le basculement

Il est recommandé d'effectuer systématiquement un basculement de test lorsqu'une modification est apportée à la configuration du stockage protégé des machines virtuelles. Ainsi, en cas d'incident, vous avez l'assurance que site Recovery Manager peut restaurer les services au sein de la cible de délai de restauration prévue.

NetApp recommande également de confirmer occasionnellement les fonctionnalités des applications chez l'invité, en particulier après la reconfiguration du stockage des machines virtuelles.

Lors de l'exécution d'une opération de restauration test, un réseau de bulles de test privé est créé sur l'hôte ESXi pour les machines virtuelles. Cependant, ce réseau n'est pas automatiquement connecté à aucune carte réseau physique et ne fournit donc pas de connectivité entre les hôtes ESXi. Pour permettre la communication entre les machines virtuelles s'exécutant sur différents hôtes ESXi lors du test de reprise après incident, un réseau privé physique est créé entre les hôtes ESXi du site de reprise après incident. Pour vérifier que le réseau de test est privé, le réseau de bulles de test peut être séparé physiquement ou à l'aide de VLAN ou de balisage VLAN. Ce réseau doit être isolé du réseau de production car les machines virtuelles sont restaurées. En effet, ils ne peuvent pas être placés sur le réseau de production avec des adresses IP qui pourraient entrer en conflit avec les systèmes de production réels. Lors de la création d'un plan de reprise d'activité dans SRM, le réseau test créé peut être sélectionné comme réseau privé afin de connecter les VM à pendant le test.

Une fois le test validé et n'est plus nécessaire, effectuez une opération de nettoyage. Le nettoyage en cours d'exécution renvoie l'état initial des machines virtuelles protégées à leur état initial et réinitialise le plan de restauration en mode prêt.

Considérations relatives au basculement

Il y a plusieurs autres considérations lorsqu'il s'agit de basculer sur un site en plus de l'ordre des opérations mentionné dans ce guide.

Vous devrez peut-être résoudre ce problème en tenant compte des différences de réseau entre les sites. Certains environnements peuvent utiliser les mêmes adresses IP réseau à la fois sur le site primaire et sur le site de reprise après incident. Cette fonctionnalité est appelée VLAN (Virtual LAN) étendu ou configuration réseau étendu. Dans d'autres environnements, il est parfois nécessaire d'utiliser différentes adresses IP réseau (par exemple, sur différents VLAN) sur le site primaire par rapport au site de reprise.

VMware offre plusieurs moyens de résoudre ce problème. Pour la première, des technologies de virtualisation de réseau comme VMware NSX-T Data Center extraient la pile réseau des couches 2 à 7 de l'environnement d'exploitation, afin d'offrir des solutions plus portables. En savoir plus sur "Options NSX-T avec SRM".

SRM vous permet également de modifier la configuration réseau d'une machine virtuelle lors de sa restauration. Cette reconfiguration inclut des paramètres tels que les adresses IP, les adresses de passerelle et les paramètres du serveur DNS. Différents paramètres réseau, qui sont appliqués aux machines virtuelles individuelles au fur et à mesure qu'elles sont restaurées, peuvent être spécifiés dans les paramètres de propriété d'une machine virtuelle dans le plan de reprise.

Pour configurer SRM de façon à appliquer différents paramètres réseau à plusieurs machines virtuelles sans devoir modifier les propriétés de chacune d'entre elles dans le plan de reprise, VMware fournit un outil appelé dr-ip-customizer. Pour savoir comment utiliser cet utilitaire, reportez-vous à la section "Documentation de VMware".

Reprotéger

Après une restauration, le site de reprise devient le nouveau site de production. Comme l'opération de reprise a rompue la réplication SnapMirror, le nouveau site de production n'est pas protégé contre un futur incident. Il est recommandé de protéger le nouveau site de production sur un autre site immédiatement après une restauration. Si le site de production d'origine est opérationnel, l'administrateur VMware peut utiliser le site de production d'origine comme nouveau site de reprise pour protéger le nouveau site de production, ce qui inversera efficacement la direction de la protection. La reprotection est disponible uniquement en cas de défaillance majeure. Par conséquent, les serveurs vCenter d'origine, les serveurs ESXi, les serveurs SRM et les bases de données correspondantes doivent être récupérables. S'ils ne sont pas disponibles, un nouveau groupe de protection et un nouveau plan de récupération doivent être créés.

Du rétablissement

Une opération de retour arrière est fondamentalement un basculement dans une direction différente de celle précédente. Il est recommandé de vérifier que le site d'origine fonctionne à un niveau de fonctionnalité acceptable avant de tenter un retour arrière ou, en d'autres termes, un basculement vers le site d'origine. Si le site d'origine est toujours compromis, vous devez reporter la restauration jusqu'à ce que la défaillance soit suffisamment remédiée.

Une autre meilleure pratique de restauration consiste à toujours effectuer un basculement de test après avoir terminé la reprotection et avant de procéder à la restauration finale. Cela vérifie que les systèmes en place sur le site initial peuvent mener à bien l'opération.

Reprotéger le site d'origine

Après la restauration, vous devez confirmer auprès de toutes les parties prenantes que leurs services ont été renvoyés à la normale avant d'exécuter à nouveau reprotéger.

La reprotection après le retour arrière reprend l'état où il était au début, avec la réplication SnapMirror à nouveau en cours d'exécution depuis le site de production vers le site de reprise.

Topologies de réplication

Dans ONTAP 9, les composants physiques d'un cluster sont visibles pour les administrateurs du cluster, mais ils ne sont pas directement visibles pour les applications et les hôtes qui utilisent le cluster. Les composants physiques offrent un pool de ressources partagées à partir duquel les ressources logiques du cluster sont créées. Les applications et les hôtes accèdent aux données uniquement au moyen de SVM qui contiennent des volumes et des LIF.

Chaque SVM NetApp est traité comme une baie dans VMware vCenter site Recovery Manager. VLSR prend en charge certaines dispositions de réplication baie à baie (ou SVM à SVM).

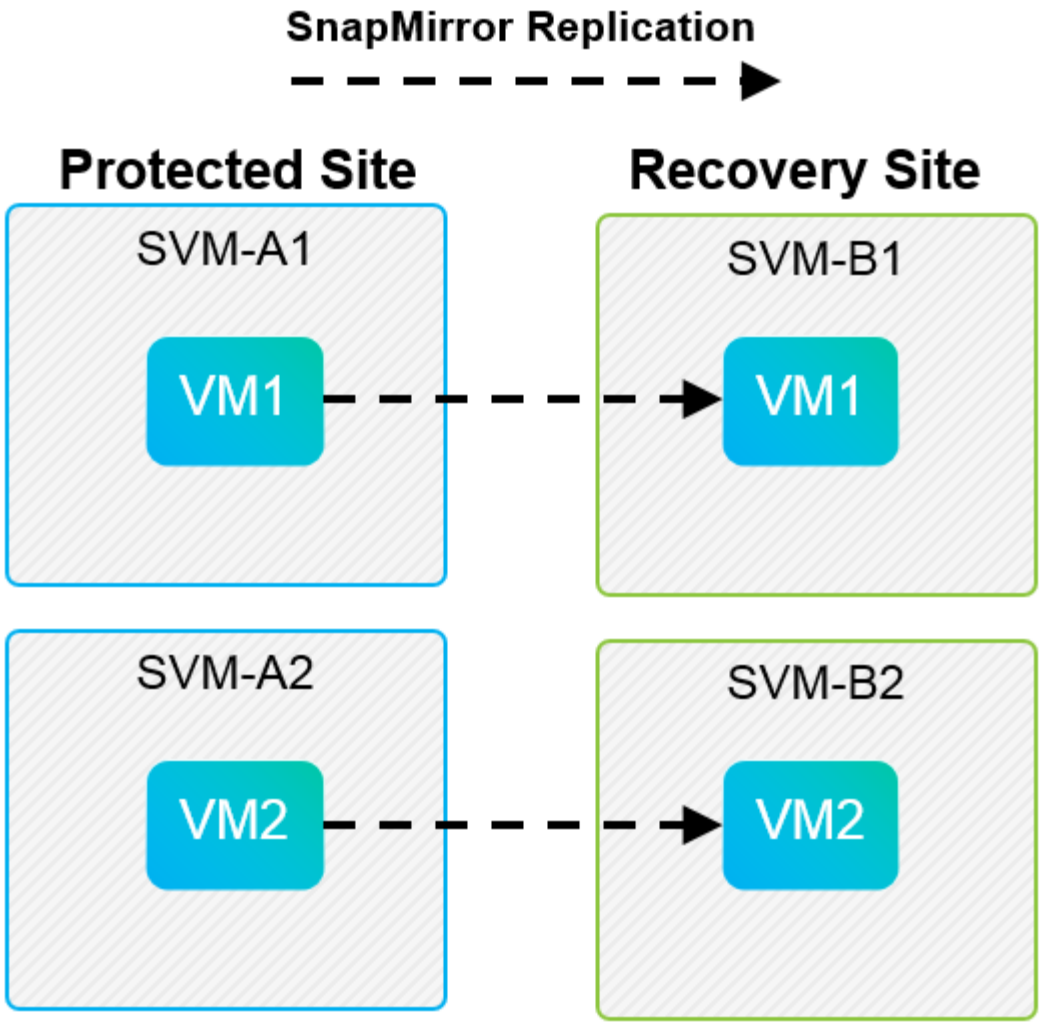
Une seule machine virtuelle ne peut pas héberger de données (Virtual machine Disk (VMDK) ou RDM) sur plusieurs baies VLSR pour les raisons suivantes :

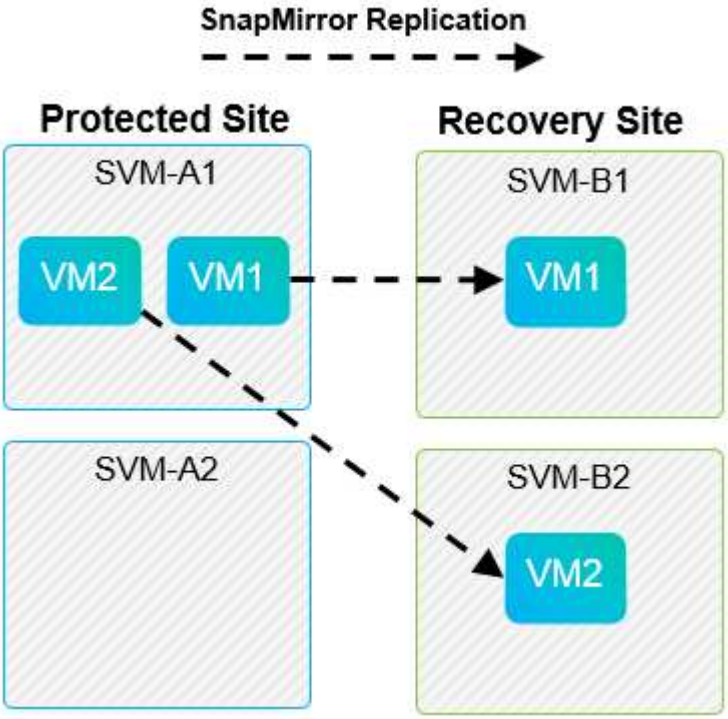
- VLSR ne voit que la SVM, pas un contrôleur physique individuel.
- Un SVM peut contrôler les LUN et les volumes répartis sur plusieurs nœuds dans un cluster.

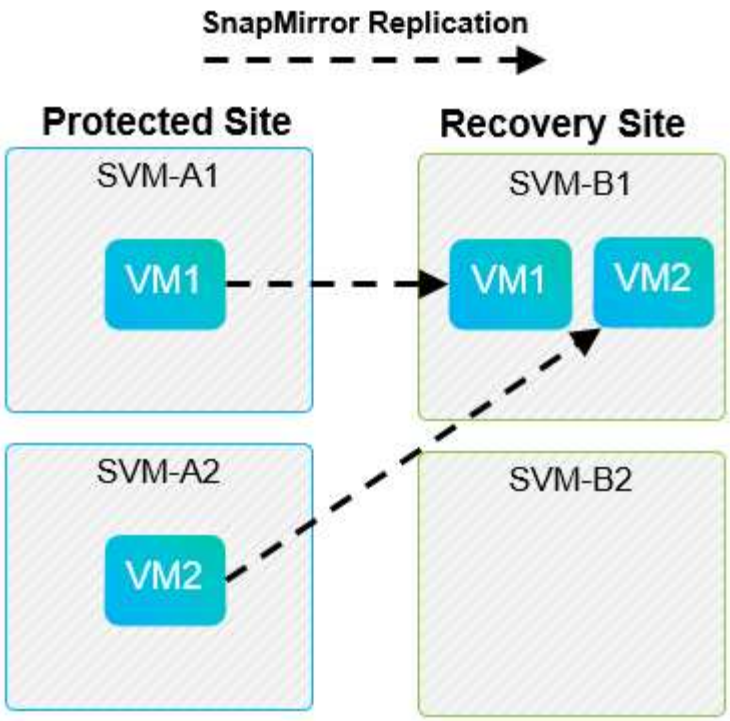
Meilleure pratique
Pour déterminer la prise en charge, conservez cette règle à l'esprit : pour protéger une machine virtuelle via VLSR et NetApp SRA, tous les composants de la machine virtuelle doivent exister sur un seul SVM. Cette règle s'applique aussi bien au site protégé que au site de reprise.

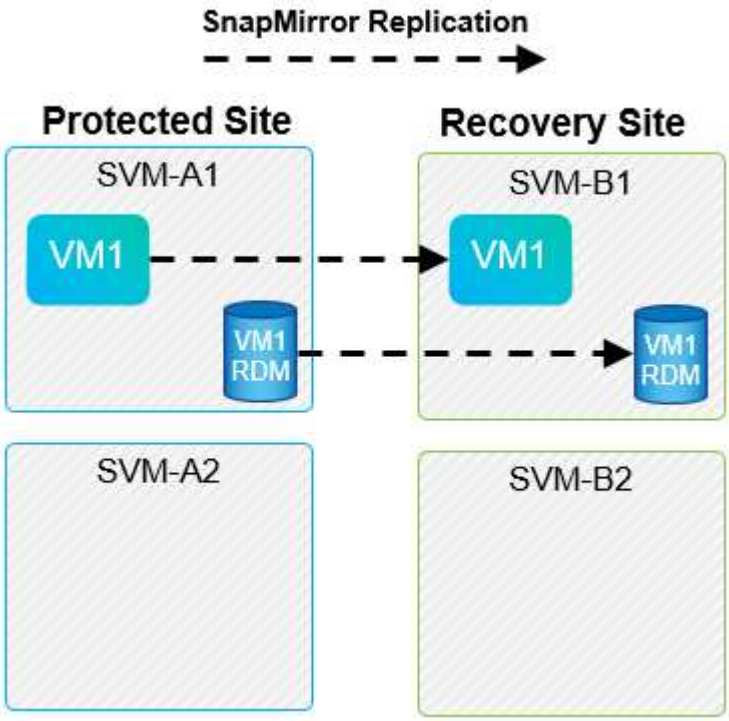
Dispositions SnapMirror prises en charge

Les figures suivantes présentent les scénarios de disposition des relations SnapMirror pris en charge par VLSR et SRA. Chaque machine virtuelle des volumes répliqués est propriétaire de données sur une seule baie VLSR (SVM) sur chaque site.


SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
VM1
SVM-B1
VM1
SVM-A2
VM2
SVM-B2
VM2

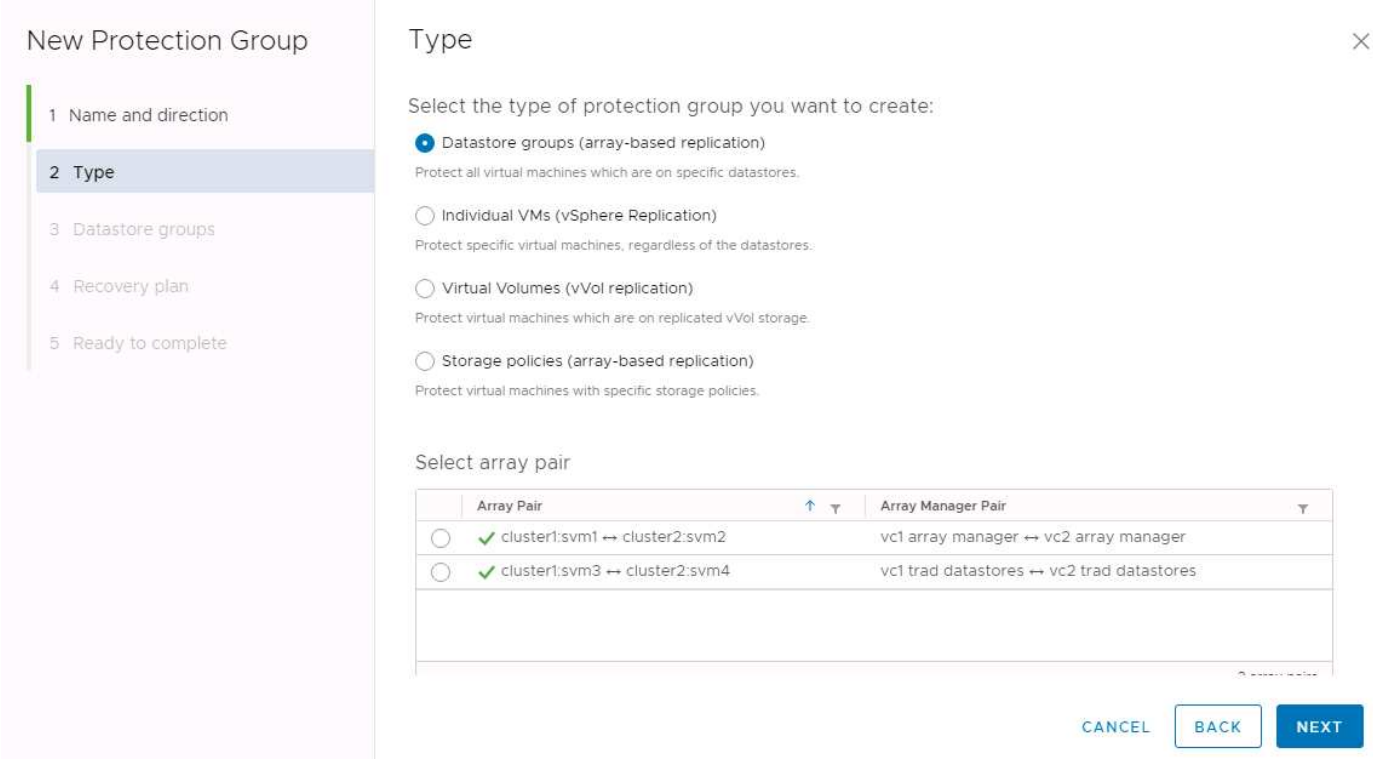

SnapMirror Replication
Protected Site
Recovery Site
SVM-A1
VM2
VM1
SVM-B1
VM1
SVM-A2
SVM-B2
VM2





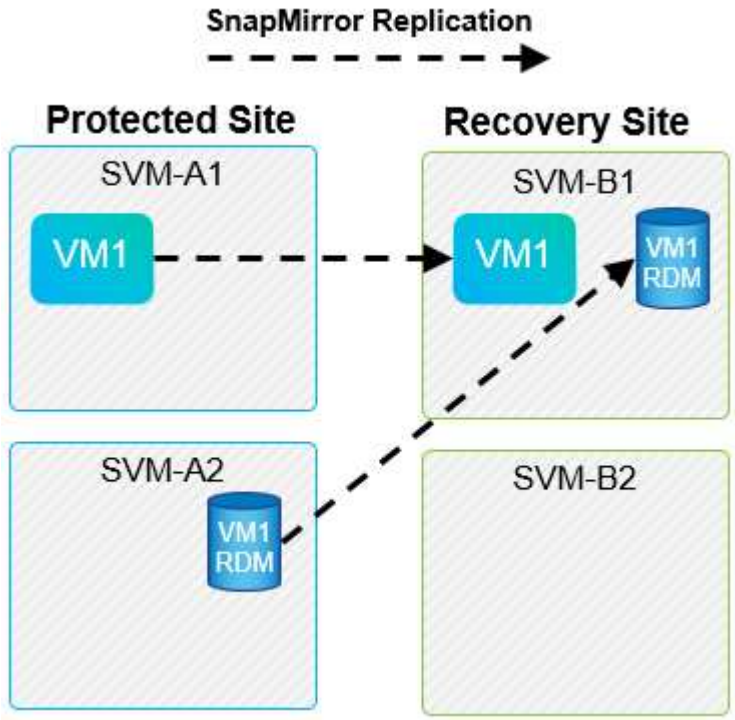
Mises en page de Array Manager prises en charge

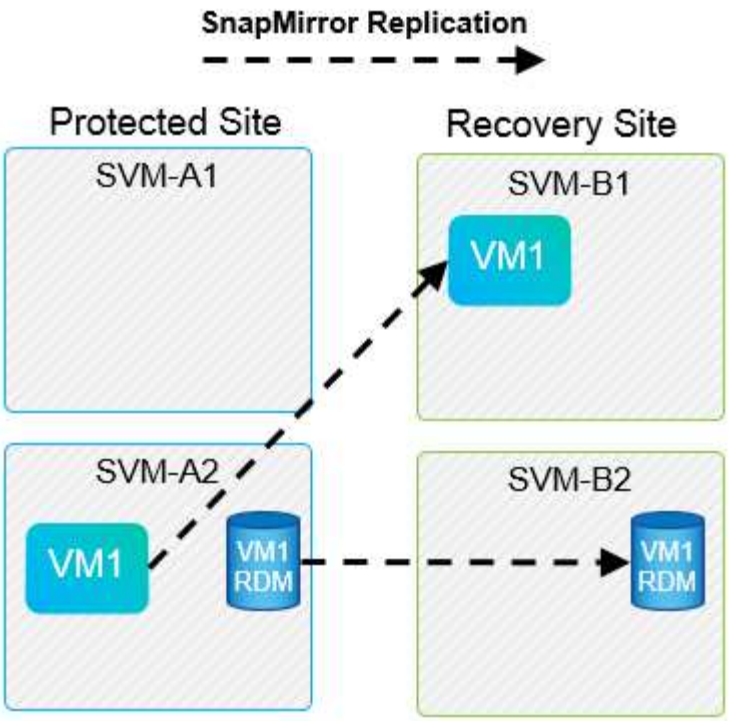
Lorsque vous utilisez la réplication basée sur la baie (ABR) dans VLSR, les groupes de protection sont isolés vers une seule paire de baies, comme l'illustre la capture d'écran suivante. Dans ce scénario, `SVM1` et `SVM2` sont associés à `SVM3` et `SVM4` sur le site de reprise. Cependant, vous ne pouvez sélectionner qu'une des deux paires de matrices lorsque vous créez un groupe de protection.



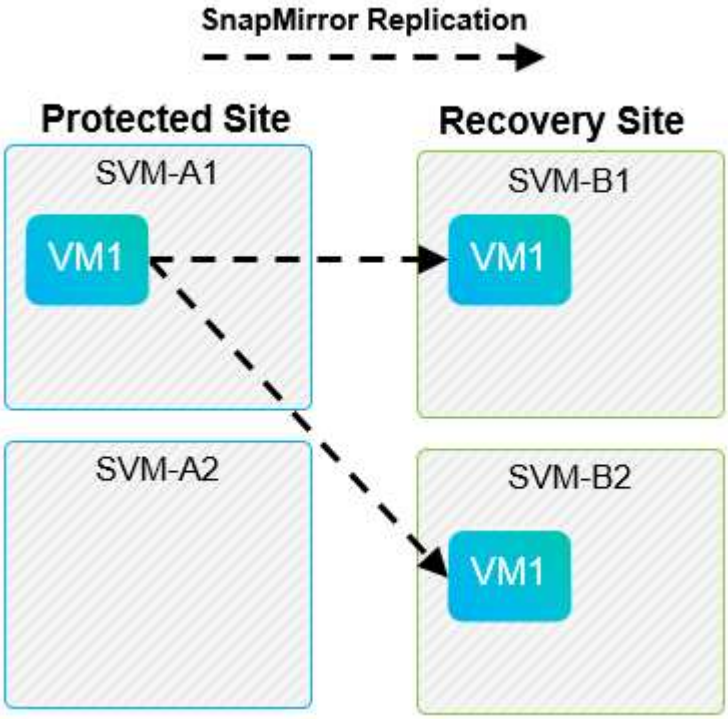
Présentations non prises en charge

Les configurations non prises en charge possèdent des données (VMDK ou RDM) sur plusieurs SVM appartenant à une machine virtuelle individuelle. Dans les exemples présentés dans les figures suivantes, `VM1` ne peut pas être configuré pour une protection avec VLSR car `VM1` possède des données sur deux SVM.





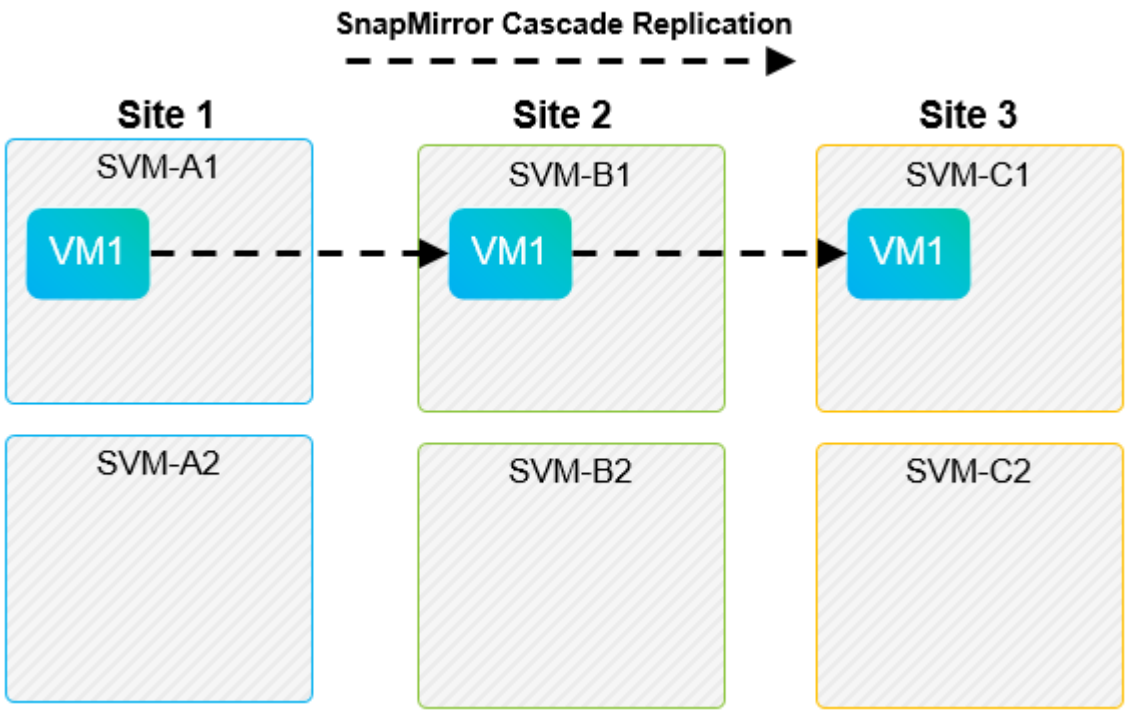
Toute relation de réplication dans laquelle un volume NetApp individuel est répliqué depuis un SVM source vers plusieurs destinations dans un même SVM ou dans différents SVM, est appelée « Fan-Out » de SnapMirror. La réplication « Fan-Out » n'est pas prise en charge par VLSR. Dans l'exemple illustré dans la figure suivante, `VM1` ne peut pas être configuré pour la protection dans VLSR car il est répliqué avec SnapMirror à deux emplacements différents.



SnapMirror en cascade

VLSR ne prend pas en charge le cascade des relations SnapMirror, dans lesquelles un volume source est répliqué sur un volume de destination, et ce volume de destination est également répliqué avec SnapMirror

vers un autre volume de destination. Dans le scénario illustré dans la figure suivante, VLSR ne peut pas être utilisé pour le basculement entre des sites.



SnapMirror et SnapVault

Le logiciel NetApp SnapVault permet de sauvegarder les données d'entreprise sur disque entre les systèmes de stockage NetApp. SnapVault et SnapMirror peuvent coexister dans un même environnement, mais VLSR prend en charge le basculement de uniquement les relations SnapMirror.

> L'adaptateur NetApp SRA prend en charge le `mirror-vault` type de règle.

SnapVault a été entièrement reconstruit pour ONTAP 8.2. Bien que les anciens utilisateurs de Data ONTAP 7-mode trouvent des similarités, des améliorations majeures ont été apportées dans cette version d'SnapVault. Une avancée majeure est la capacité à préserver l'efficacité du stockage sur les données primaires au cours des transferts SnapVault.

L'architecture SnapVault de ONTAP 9 réplique au niveau du volume et non au niveau du qtree, comme c'est le cas avec 7-mode SnapVault. Dans ce cas, la source d'une relation SnapVault doit être un volume, et ce volume doit être répliqué sur son propre volume sur le système secondaire SnapVault.

Dans un environnement dans lequel SnapVault est utilisé, des snapshots nommés spécifiques sont créés sur le système de stockage principal. Selon la configuration implémentée, les snapshots nommés peuvent être créés sur le système principal par une planification SnapVault ou par une application telle que NetApp Active IQ Unified Manager. Les snapshots nommés créés sur le système primaire sont ensuite répliqués sur la destination SnapMirror, puis stockés sur la destination SnapVault.

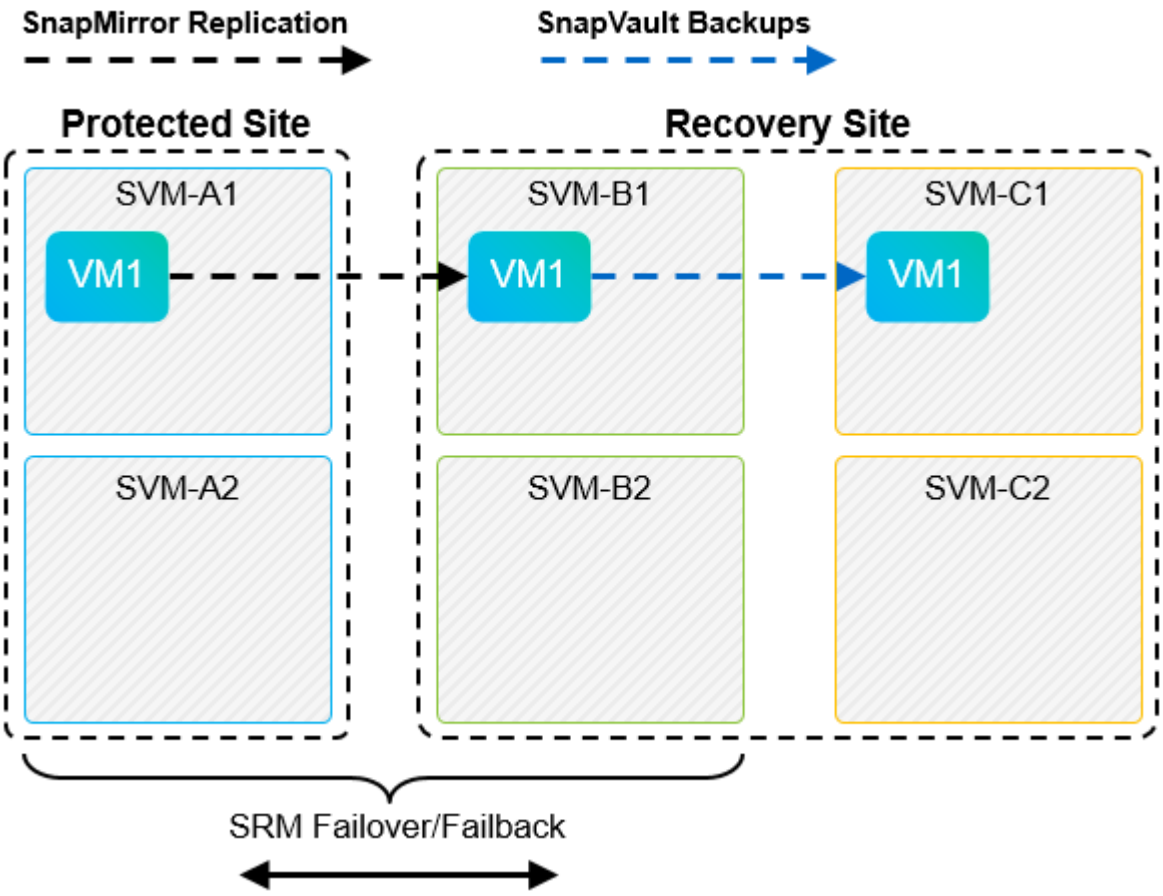
Un volume source peut être créé dans une configuration en cascade, dans laquelle un volume est répliqué vers une destination SnapMirror dans le site de reprise après incident, et depuis ce volume est copié vers une destination SnapVault. Un volume source peut également être créé au sein d'une relation « fan-out » où une destination est une destination SnapMirror et l'autre destination est une destination SnapVault. Toutefois, SRA ne reconfigurez pas automatiquement la relation SnapVault pour utiliser le volume de destination SnapMirror comme source du coffre-fort en cas de basculement ou d'inversion de réplication VLSR.

Pour connaître les dernières informations concernant SnapMirror et SnapVault pour ONTAP 9, consultez "Tr-4015 Guide des meilleures pratiques en matière de configuration de SnapMirror pour ONTAP 9."

Meilleure pratique
Si SnapVault et VLSR sont utilisés dans le même environnement, NetApp recommande d'utiliser une configuration SnapMirror vers SnapVault en cascade dans laquelle les sauvegardes SnapVault sont normalement exécutées à partir de la destination SnapMirror sur le site de reprise après incident. En cas d'incident, cette configuration rend le site principal inaccessible. Le fait de conserver la destination SnapVault sur le site de reprise permet de reconfigurer les sauvegardes SnapVault après le basculement, de sorte que les sauvegardes SnapVault puissent continuer sur le site de reprise.

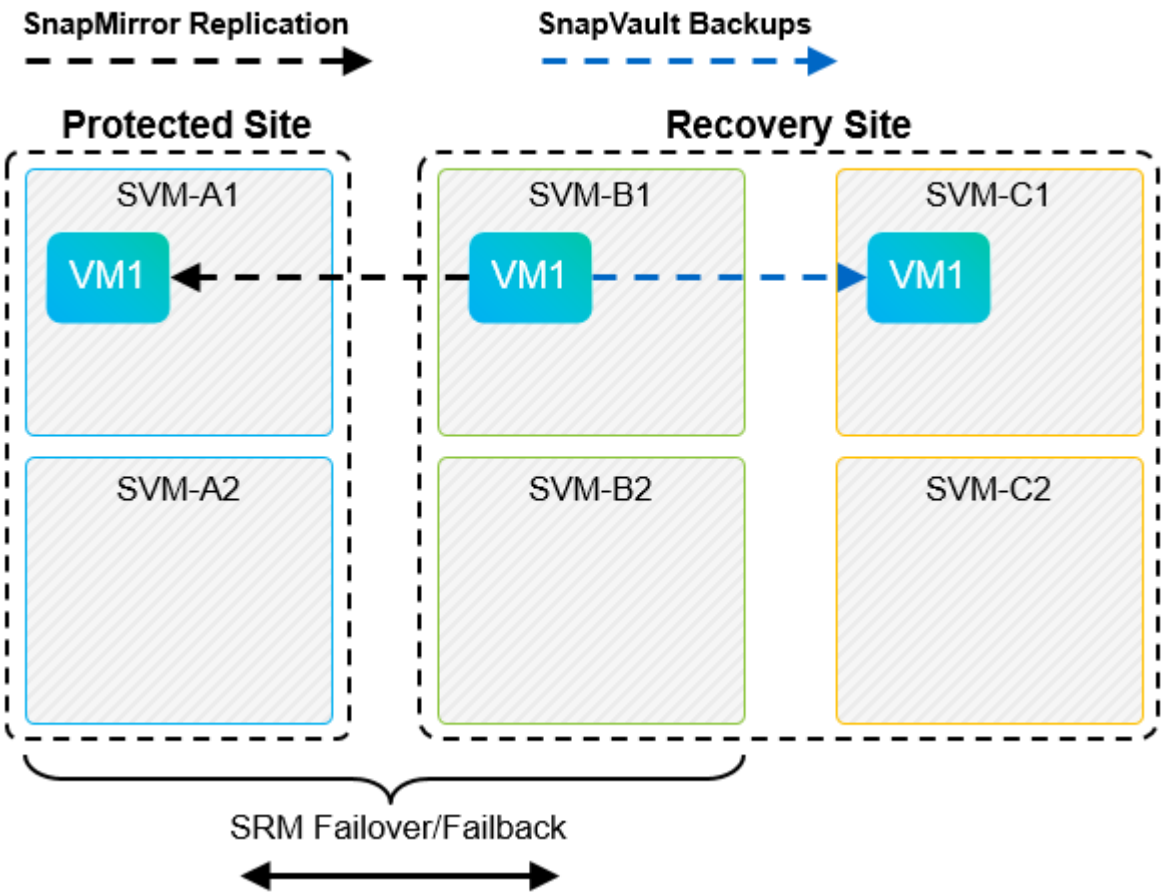
Dans un environnement VMware, chaque datastore dispose d'un identifiant unique universel (UUID) et chaque machine virtuelle possède un ID d'objet géré unique (MOID). Ces identifiants ne sont pas gérés par VLSR lors du basculement ou de la restauration. Étant donné que les UID et les MOID de machine virtuelle ne sont pas maintenus lors du basculement par VLSR, toutes les applications qui dépendent de ces ID doivent être reconfigurées après le basculement VLSR. NetApp Active IQ Unified Manager, qui coordonne la réplication SnapVault avec l'environnement vSphere, est un exemple d'application.

La figure suivante décrit une configuration SnapMirror vers SnapVault en cascade. Si la destination SnapVault se trouve sur le site de reprise après incident ou sur un site tertiaire non affecté par une panne sur le site primaire, l'environnement peut être reconfiguré afin de permettre la continuité des sauvegardes après le basculement.



La figure suivante décrit la configuration après l'utilisation de VLSR pour renvoyer la réplication SnapMirror vers le site primaire. L'environnement a également été reconfiguré de façon à ce que les sauvegardes SnapVault s'effectuent à partir d'une source SnapMirror. Cette configuration est « Fan-Out » de SnapMirror

SnapVault.



Une fois que vsrm a effectué le retour arrière et une seconde inversion des relations SnapMirror, les données de production sont de nouveau sur le site principal. Ces données sont désormais protégées de la même manière qu'avant le basculement vers le site de reprise après incident, via les sauvegardes SnapMirror et SnapVault.

Utilisation de qtrees dans les environnements site Recovery Manager

Les qtrees sont des répertoires spéciaux qui permettent l'application de quotas de système de fichiers pour NAS. ONTAP 9 permet la création de qtrees et peut exister dans les volumes répliqués avec SnapMirror. Toutefois, SnapMirror ne permet pas la réplication de qtrees individuels ni de réplication au niveau qtree. Toute la réplication SnapMirror se fait au niveau du volume uniquement. C'est pour cette raison que NetApp ne recommande pas l'utilisation de qtrees avec VLSR.

Environnements FC et iSCSI mixtes

Grâce à la prise en charge des protocoles SAN (FC, FCoE et iSCSI), ONTAP 9 propose des services LUN, à savoir la création de LUN et leur mappage vers les hôtes associés. Dans la mesure où le cluster compte plusieurs contrôleurs, il existe plusieurs chemins logiques gérés par les E/S multivoies vers une LUN individuelle. L'accès ALUA (Asymmetric Logical Unit Access) est utilisé sur les hôtes pour que le chemin optimisé vers un LUN soit sélectionné et activé pour le transfert de données. Si ce chemin change (par exemple, en raison du déplacement du volume qui y est associé), ONTAP 9 reconnaît automatiquement cette modification et s'ajuste de façon non disruptive. S'il devient indisponible, ONTAP peut également basculer sans interruption sur un autre chemin.

VMware VLSR et NetApp SRA prennent en charge l'utilisation du protocole FC sur un site et le protocole iSCSI

sur l'autre site. Il ne prend pas en charge la combinaison de datastores FC et de datastores iSCSI dans le même hôte ESXi ou d'hôtes différents dans le même cluster. Cette configuration n'est pas prise en charge avec VLSR car, pendant le basculement VLSR ou le basculement de test, VLSR inclut tous les initiateurs FC et iSCSI des hôtes ESXi dans la demande.

Meilleure pratique
VLSR et SRA prennent en charge les protocoles FC et iSCSI mixtes entre les sites protégés et de reprise. Cependant, chaque site ne doit pas être configuré avec un seul protocole, FC ou iSCSI, et non avec les deux protocoles sur le même site. Si il est nécessaire de configurer les protocoles FC et iSCSI sur le même site, NetApp recommande que certains hôtes utilisent iSCSI et d'autres hôtes utilisent FC. Dans ce cas, NetApp recommande également de configurer les mappages de ressources VLSR de sorte que les VM soient configurés pour basculer vers un groupe d'hôtes ou un autre.

Dépannage de VLSRM/SRM lors de l'utilisation de la réplication vVols

Lors de l'utilisation des outils ONTAP 9.13P2, le workflow au sein de VLSR et de SRM est très différent lors de l'utilisation de la réplication vVols par rapport à l'utilisation de SRA et des datastores traditionnels. Par exemple, il n'existe pas de concept de gestionnaire de baie. Ainsi, `discoverarrays` les commandes et `discoverdevices` ne sont jamais vues.

Lors du dépannage, il est utile de comprendre les nouveaux flux de travail répertoriés ci-dessous :

1. QueryReplicationPeer : détecte les accords de réplication entre deux domaines de défaillance.
2. QueryFaultDomain : détecte la hiérarchie du domaine de pannes.
3. QueryReplicationGroup : détecte les groupes de réplication présents dans les domaines source ou cible.
4. SyncReplicationGroup : synchronise les données entre la source et la cible.
5. QueryPointInTimeReplica : détecte le point dans le temps des répliques sur une cible.
6. TestFailoverReplicationGroupStart : démarre le basculement de test.
7. TestFailoverReplicationGroupStop : met fin au basculement de test.
8. PromoteReplicationGroup : promeut un groupe actuellement en cours de test à la production.
9. PreparFailoverReplicationTM : prépare une reprise après sinistre.
10. FailoverReplicationGroup : exécute la reprise après incident.
11. ReverseReplicateGroup : lance la réplication inverse.
12. QueryMatchingContainer : recherche les conteneurs (ainsi que les hôtes ou les groupes de réplication) susceptibles de satisfaire une demande de provisionnement avec une règle donnée.
13. QueryResourceMetadata : recherche les métadonnées de toutes les ressources du fournisseur VASA, l'utilisation des ressources peut être renvoyée comme réponse à la fonction queryMatchingContainer.

L'erreur la plus courante lors de la configuration de la réplication vvols est une incapacité à découvrir les relations SnapMirror. En effet, les volumes et les relations SnapMirror sont créés en dehors de la purView des outils ONTAP. Il est donc recommandé de toujours s'assurer que votre relation SnapMirror est totalement initialisée et que vous avez exécuté une redécouverte dans les outils ONTAP sur les deux sites avant de tenter de créer un datastore vvols répliqué.

Informations supplémentaires

Pour en savoir plus sur les informations données dans ce livre blanc, consultez ces documents et/ou sites web :

- Ressources relatives aux outils ONTAP pour VMware vSphere 10.x. "https://mysupport.netapp.com/site/products/all/details/otv10/docs-tab"
- Ressources relatives aux outils ONTAP pour VMware vSphere 9.x. "https://mysupport.netapp.com/site/products/all/details/otv/docsandkb-tab"
- Tr-4597 : VMware vSphere pour ONTAP "https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vsphere-overview.html"
- Tr-4400 : volumes virtuels VMware vSphere avec ONTAP "https://docs.netapp.com/us-en/ontap-apps-dbs/vmware/vmware-vvols-overview.html"
- Tr-4015 Guide des meilleures pratiques en matière de configuration de SnapMirror pour ONTAP 9 https://www.netapp.com/media/17229-tr4015.pdf?v=127202175503P
- Documentation de VMware Live site Recovery "https://techdocs.broadcom.com/us/en/vmware-cis/live-recovery/live-site-recovery/9-0.html"

Reportez-vous à la "Matrice d'interopérabilité (IMT)"sur le site de support NetApp pour vous assurer que les versions de produits et de fonctionnalités mentionnées dans le présent document sont prises en charge par votre environnement. NetApp IMT définit les composants et versions de produits qu'il est possible d'utiliser pour créer des configurations prises en charge par NetApp. Les résultats dépendent des installations de chaque client et de leur conformité aux spécifications publiées.

Cluster de stockage vSphere Metro avec ONTAP

Cluster de stockage vSphere Metro avec ONTAP

L'hyperviseur vSphere de pointe de VMware peut être déployé en tant que cluster étendu appelé vMSC (vSphere Metro Storage Cluster).

Les solutions VMSC sont prises en charge avec NetApp® MetroCluster™ et la synchronisation active SnapMirror (anciennement appelée SnapMirror Business Continuity ou SMBC) et assurent une continuité de l'activité avancée si un ou plusieurs domaines à défaillance subissent une panne totale. La résilience aux différents modes de défaillance dépend des options de configuration que vous choisissez.

> Cette documentation remplace les rapports techniques *TR-4128: VSphere on NetApp MetroCluster*

Disponibilité continue pour les environnements vSphere

L'architecture ONTAP est une plateforme de stockage flexible et évolutive qui fournit des services SAN (FCP, iSCSI et NVMe-of) et NAS (NFS v3 et v4.1) pour les datastores. Les systèmes de stockage NetApp AFF, ASA et FAS utilisent le système d'exploitation ONTAP pour offrir des protocoles supplémentaires pour l'accès au stockage invité comme S3 et SMB/CIFS.

NetApp MetroCluster utilise la fonction HA (basculement du contrôleur ou CFO) de NetApp pour se protéger contre les défaillances du contrôleur. Elle comprend également la technologie SyncMirror locale, le basculement de cluster en cas d'incident (basculement de cluster en cas d'incident ou CFOD), la redondance matérielle et la séparation géographique pour atteindre des niveaux de disponibilité élevés. SyncMirror met en

miroir les données de manière synchrone sur les deux moitiés de la configuration MetroCluster en écrivant les données sur deux plexes : le plex local (sur le tiroir local) assure activement le service des données et le plex distant (sur le tiroir distant) n'assure généralement pas le service des données. La redondance matérielle est mise en place pour tous les composants MetroCluster, tels que les contrôleurs, le stockage, les câbles, les commutateurs (utilisés avec Fabric MetroCluster) et les adaptateurs.

La synchronisation active NetApp SnapMirror, disponible sur les systèmes non MetroCluster et ASA r2, offre une protection granulaire du datastore avec les protocoles SAN FCP et iSCSI. Il vous permet soit de protéger l'intégralité du vMSC, soit de protéger de manière sélective les workloads prioritaires. Il offre un accès actif/actif aux sites locaux et distants, contrairement à NetApp MetroCluster, qui est une solution de secours actif. Depuis la version ONTAP 9.15.1, la synchronisation active SnapMirror prend en charge une fonctionnalité actif-actif symétrique. Elle permet d'effectuer des opérations de lecture et d'écriture d'E/S à partir des deux copies d'un LUN protégé grâce à une réplication synchrone bidirectionnelle, ce qui permet aux deux copies de LUN de traiter les opérations d'E/S localement. Avant ONTAP 9.15.1, la synchronisation active SnapMirror ne prend en charge que les configurations actives/actives asymétriques, dans lesquelles les données du site secondaire sont proxys sur la copie principale d'un LUN.

Pour créer un cluster VMware HA/DRS sur deux sites, les hôtes ESXi sont utilisés et gérés par une appliance vCenter Server (VCSA). Les réseaux de gestion vSphere, vMotion® et machine virtuelle sont connectés via un réseau redondant entre les deux sites. Le serveur vCenter gérant le cluster HA/DRS peut se connecter aux hôtes ESXi sur les deux sites et doit être configuré à l'aide de vCenter HA.

Reportez-vous à la section "Comment créer et configurer des clusters dans le client vSphere" Pour configurer vCenter HA.

Vous devez également vous reporter "Bonnes pratiques pour VMware vSphere Metro Storage Cluster" à .

Qu'est-ce que le cluster de stockage vSphere Metro ?

VSphere Metro Storage Cluster (vMSC) est une configuration certifiée qui protège les machines virtuelles et les conteneurs contre les défaillances. Pour y parvenir, les concepts de stockage étendus ainsi que les clusters d'hôtes ESXi sont répartis sur différents domaines à défaillance, tels que les racks, les bâtiments, les campus ou même les villes. Les technologies de stockage Active Sync de NetApp MetroCluster et SnapMirror assurent une protection avec un objectif de point de récupération nul (RPO=0) aux clusters hôtes. La configuration vMSC est conçue pour garantir la disponibilité continue des données, même en cas de défaillance d'un « site » physique ou logique complet. Un périphérique de stockage faisant partie de la configuration vMSC doit être certifié après avoir suivi un processus de certification vMSC réussi. Tous les périphériques de stockage pris en charge se trouvent dans le "Guide de compatibilité du stockage VMware".

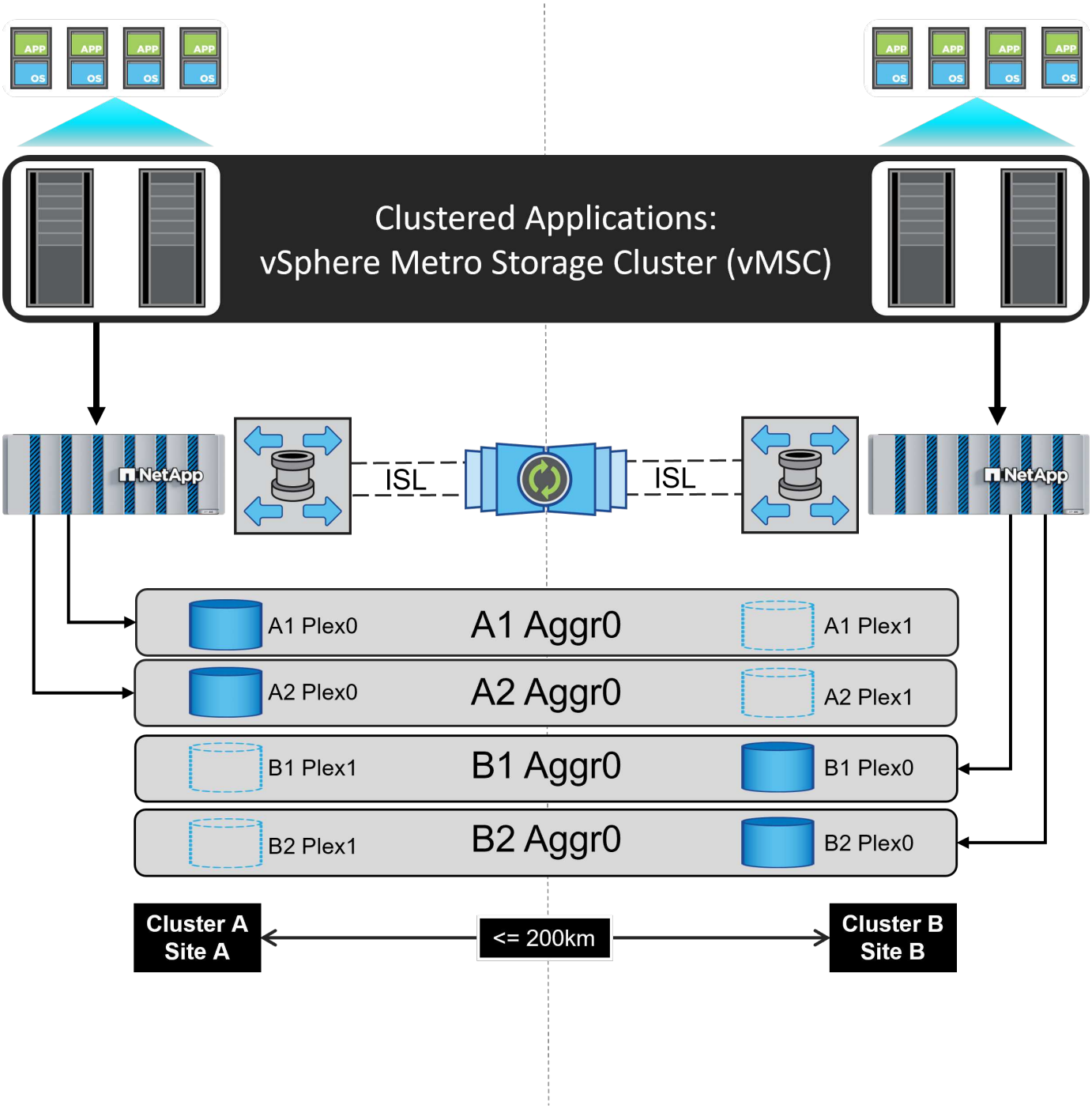
Pour plus d'informations sur les conseils de conception pour vSphere Metro Storage Cluster, reportez-vous à la documentation suivante :

- "Prise en charge de VMware vSphere avec NetApp MetroCluster"
- "Prise en charge de VMware vSphere avec la continuité de l'activité NetApp SnapMirror" (Maintenant appelé synchronisation active SnapMirror)

NetApp MetroCluster peut être déployé dans deux configurations différentes pour une utilisation avec vSphere :

- MetroCluster extensible
- MetroCluster de structure

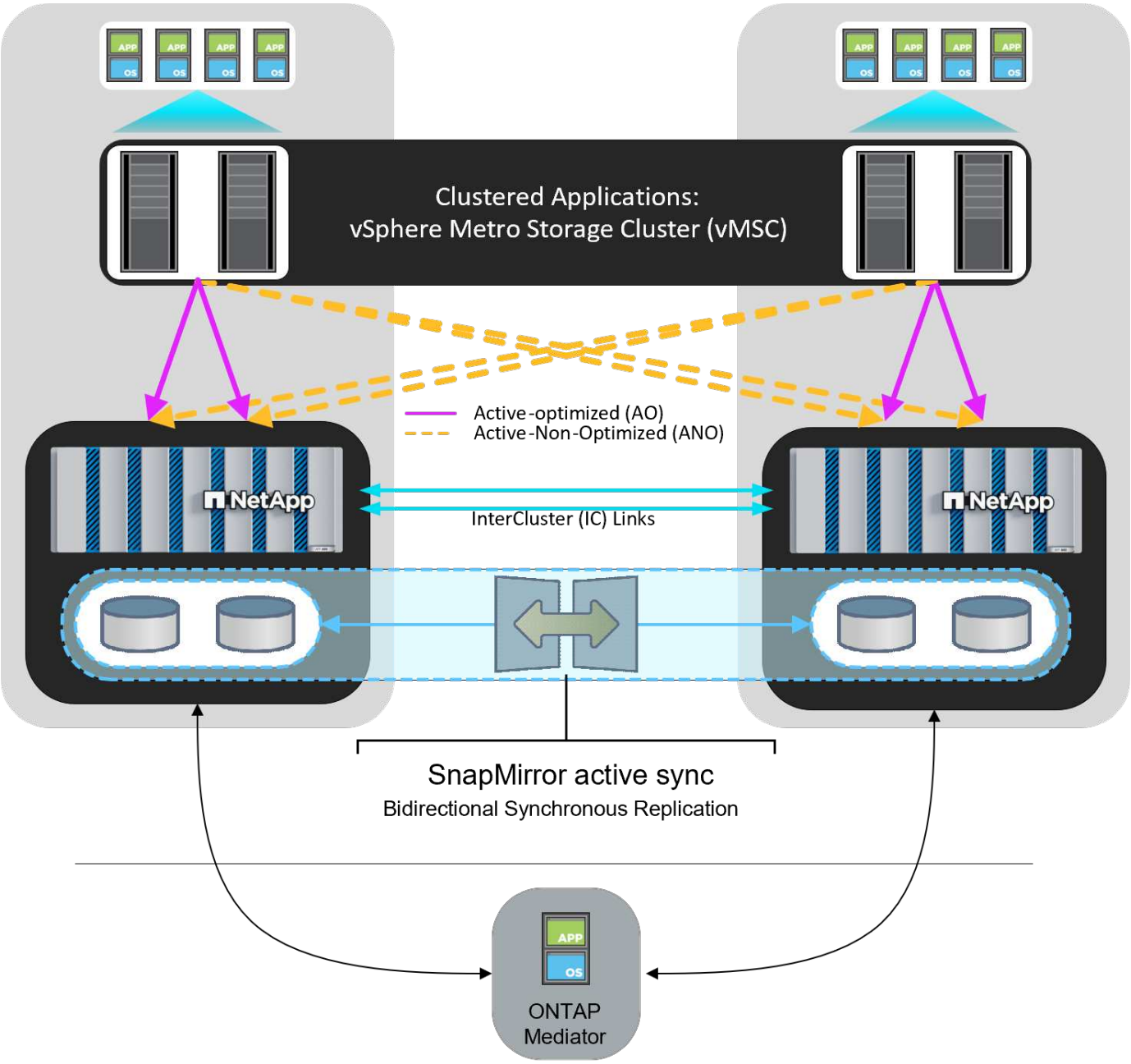
Voici une illustration de la topologie générale d'Stretch MetroCluster.



Reportez-vous à la section "Documentation MetroCluster" Pour obtenir des informations spécifiques sur la conception et le déploiement de MetroCluster.

La synchronisation active SnapMirror peut également être déployée de deux manières différentes.

- Asymétrique
- Synchronisation active symétrique (ONTAP 9.15.1)



Reportez-vous "Documents NetApp" à la pour obtenir des informations spécifiques sur la conception et le déploiement de la synchronisation active SnapMirror.

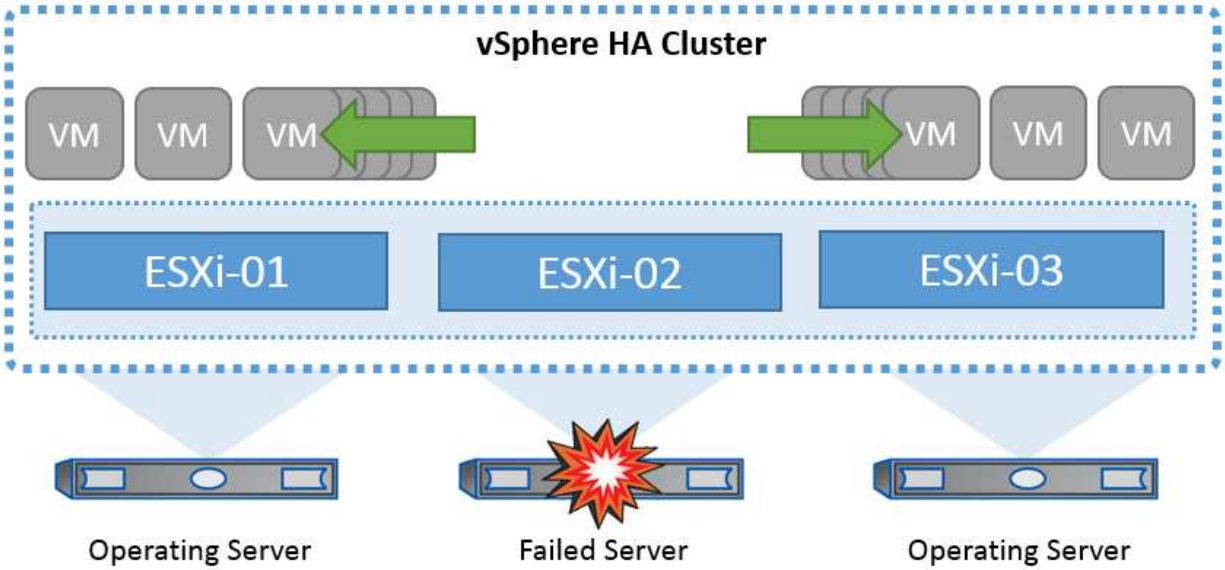
Présentation de la solution VMware vSphere

VCenter Server Appliance (VCSA) est le puissant système de gestion centralisée et une interface unique pour vSphere qui permet aux administrateurs d'exploiter efficacement les clusters ESXi. Cet outil facilite les fonctions clés telles que le provisionnement des machines virtuelles, les opérations vMotion, la haute disponibilité (HA), Distributed Resource Scheduler (DRS), Tanzu Kubernetes Grid et bien plus encore. Elle constitue un composant essentiel des environnements clouds VMware et doit être conçue en tenant compte de la disponibilité du service.

Haute disponibilité vSphere

La technologie de cluster de VMware regroupe les serveurs ESXi en pools de ressources partagées pour les machines virtuelles et fournit la haute disponibilité vSphere (HA). VSphere HA offre une haute disponibilité facile à utiliser pour les applications exécutées sur les machines virtuelles. Lorsque la fonction de haute disponibilité est activée sur le cluster, chaque serveur ESXi maintient la communication avec les autres hôtes de sorte que si un hôte ESXi ne répond plus ou est isolé, le cluster de haute disponibilité peut négocier la restauration des machines virtuelles qui s'exécutaient sur cet hôte ESXi parmi les hôtes survivants du cluster. En cas de panne du système d'exploitation invité, vSphere HA peut redémarrer l'ordinateur virtuel affecté sur le même serveur physique. VSphere HA permet de réduire les temps d'indisponibilité planifiés, d'éviter les temps d'indisponibilité non planifiés et de restaurer rapidement les données en cas de panne.

Cluster vSphere HA qui récupère les machines virtuelles à partir d'un serveur défaillant.



Il est important de comprendre que VMware vSphere ne connaît pas la synchronisation active NetApp MetroCluster ou SnapMirror et que tous les hôtes ESXi du cluster vSphere sont des hôtes éligibles pour les opérations de cluster haute disponibilité selon les configurations d'affinité des groupes de machines virtuelles et des hôtes.

Détection de défaillance de l'hôte

Dès la création du cluster HA, tous les hôtes du cluster participent à des élections et l'un des hôtes devient maître. Chaque esclave exécute une pulsation réseau vers le maître, et le maître effectue à son tour une pulsation réseau sur tous les hôtes esclaves. L'hôte maître d'un cluster vSphere HA est responsable de la détection de la défaillance des hôtes esclaves.

En fonction du type de défaillance détecté, les machines virtuelles exécutées sur les hôtes peuvent avoir besoin d'être basculées.

Dans un cluster vSphere HA, trois types de défaillance d'hôte sont détectés :

- Défaillance - Un hôte cesse de fonctionner.
- Isolation - Un hôte devient isolé du réseau.
- Partition : Un hôte perd la connectivité réseau avec l'hôte maître.

L'hôte maître surveille les hôtes esclaves du cluster. Cette communication s'effectue par échange de

battements de cœur réseau toutes les secondes. Lorsque l'hôte maître cesse de recevoir ces battements de cœur d'un hôte esclave, il vérifie la livenité de l'hôte avant de déclarer l'échec de l'hôte. La vérification de la livenité effectuée par l'hôte maître consiste à déterminer si l'hôte esclave échange des pulsations avec l'un des datastores. En outre, l'hôte maître vérifie si l'hôte répond aux requêtes ping ICMP envoyées à ses adresses IP de gestion pour détecter s'il est simplement isolé de son nœud maître ou complètement isolé du réseau. Pour ce faire, il exécute une commande ping sur la passerelle par défaut. Une ou plusieurs adresses d'isolement peuvent être spécifiées manuellement pour améliorer la fiabilité de la validation de l'isolement.

> NetApp recommande de spécifier au moins deux adresses d'isolement supplémentaires, et que chacune de ces adresses soit site-local. Cela améliorera la fiabilité de la validation de l'isolement.

Réponse d'isolation de l'hôte

Isolation Response est un paramètre de vSphere HA qui détermine l'action déclenchée sur les machines virtuelles lorsqu'un hôte d'un cluster vSphere HA perd ses connexions réseau de gestion mais continue à s'exécuter. Il existe trois options pour ce paramètre, « Désactivé », « Arrêter et redémarrer les machines virtuelles » et « Arrêter et redémarrer les machines virtuelles ».

« Arrêter » est préférable à « éteindre », qui ne vide pas les modifications les plus récentes apportées au disque ou ne valide pas les transactions. Si les machines virtuelles ne s'arrêtent pas dans les 300 secondes, elles sont éteintes. Pour modifier le temps d'attente, utilisez l'option avancée das.isolashutdowntimeout.

Avant que la haute disponibilité ne lance la réponse d'isolation, elle vérifie d'abord si l'agent principal vSphere HA possède le datastore qui contient les fichiers de configuration de la machine virtuelle. Si ce n'est pas le cas, l'hôte ne déclenchera pas la réponse d'isolation, car il n'y a pas de maître pour redémarrer les machines virtuelles. L'hôte vérifie régulièrement l'état du datastore pour déterminer s'il est demandé par un agent vSphere HA qui détient le rôle principal.

> NetApp recommande de définir la « réponse d'isolation de l'hôte » sur Désactivé.

Une condition de split-brain peut se produire si un hôte est isolé ou partitionné à partir de l'hôte maître vSphere HA et que le maître ne peut pas communiquer via des datastores heartbeat ou par ping. Le maître déclare l'hôte isolé comme étant mort et redémarre les machines virtuelles sur les autres hôtes du cluster. Une condition de split-brain existe maintenant parce qu'il y a deux instances de la machine virtuelle en cours d'exécution, dont une seule peut lire ou écrire les disques virtuels. Il est désormais possible d'éviter les conditions de split-brain en configurant VM Component protection (VMCP).

Protection des composants VM (VMCP)

L'une des améliorations de vSphere 6, concernant la haute disponibilité, est VMCP. VMCP offre une protection améliorée contre les conditions de tous les chemins d'accès (APD) et de perte permanente de périphérique (PDL) pour le stockage bloc (FC, iSCSI, FCoE) et de fichiers (NFS).

Perte permanente de périphérique (PDL)

PDL est une condition qui se produit lorsqu'un périphérique de stockage tombe en panne de manière permanente ou est supprimé administrativement et ne devrait pas revenir. La baie de stockage NetApp émet un code de détection SCSI pour ESXi déclarant que le périphérique est définitivement perdu. Dans la section Conditions de défaillance et réponse de la machine virtuelle de vSphere HA, vous pouvez configurer la réponse après la détection d'une condition PDL.

NetApp recommande de définir la "réponse pour le datastore avec PDL" sur "**éteindre et redémarrer les machines virtuelles**". Lorsque cette condition est détectée, une machine virtuelle est redémarrée instantanément sur un hôte sain dans le cluster vSphere HA.

Tous les chemins en panne (APD)

L'APD est une condition qui se produit lorsqu'un périphérique de stockage devient inaccessible à l'hôte et qu'aucun chemin vers la matrice n'est disponible. ESXi considère cela comme un problème temporaire avec le périphérique et s'attend à ce qu'il redevienne disponible.

Lorsqu'une condition APD est détectée, une minuterie démarre. Au bout de 140 secondes, la condition APD est officiellement déclarée et le périphérique est marqué comme étant hors délai APD. Lorsque les 140 secondes sont écoulées, la haute disponibilité commence à compter le nombre de minutes spécifié dans le délai d'attente pour le basculement de machine virtuelle. Une fois le délai spécifié écoulé, la haute disponibilité redémarre les machines virtuelles impactées. Vous pouvez configurer VMCP pour qu'il réponde différemment si vous le souhaitez (désactivé, événements de problème ou mise hors tension et redémarrage des machines virtuelles).

- NetApp recommande de configurer la "réponse pour le datastore avec APD" sur "**éteindre et redémarrer les machines virtuelles (conservative)**".
- Conservateur fait référence à la probabilité que la haute disponibilité soit capable de redémarrer les machines virtuelles. Si elle est définie sur conservateur, la haute disponibilité ne redémarrera la machine virtuelle concernée par l'APD que si elle sait qu'un autre hôte peut la redémarrer. Dans le cas d'un environnement agressif, la haute disponibilité essaiera de redémarrer la machine virtuelle même si elle ne connaît pas l'état des autres hôtes. Cela peut entraîner le redémarrage des machines virtuelles si aucun hôte n'a accès au datastore sur lequel elles se trouvent.
- Si l'état APD est résolu et que l'accès au stockage est restauré avant le délai d'expiration, HA ne redémarrera pas inutilement la machine virtuelle sauf si vous la configurez explicitement pour le faire. Si une réponse est souhaitée, même lorsque l'environnement a récupéré de la condition APD, la réponse pour la restauration APD après le délai APD doit être configurée pour réinitialiser les machines virtuelles.
- NetApp recommande de configurer la réponse pour la récupération APD après le délai APD sur Désactivé.

Implémentation de VMware DRS pour NetApp SnapMirror Active Sync

VMware DRS est une fonctionnalité qui regroupe les ressources hôtes dans un cluster et est principalement utilisée pour équilibrer la charge au sein d'un cluster dans une infrastructure virtuelle. VMware DRS calcule principalement les ressources CPU et mémoire pour effectuer l'équilibrage de charge dans un cluster. Étant donné que vSphere ne connaît pas la mise en cluster étendue, il prend en compte tous les hôtes des deux sites lors de l'équilibrage de charge.

Implémentation de VMware DRS pour NetApp MetroCluster

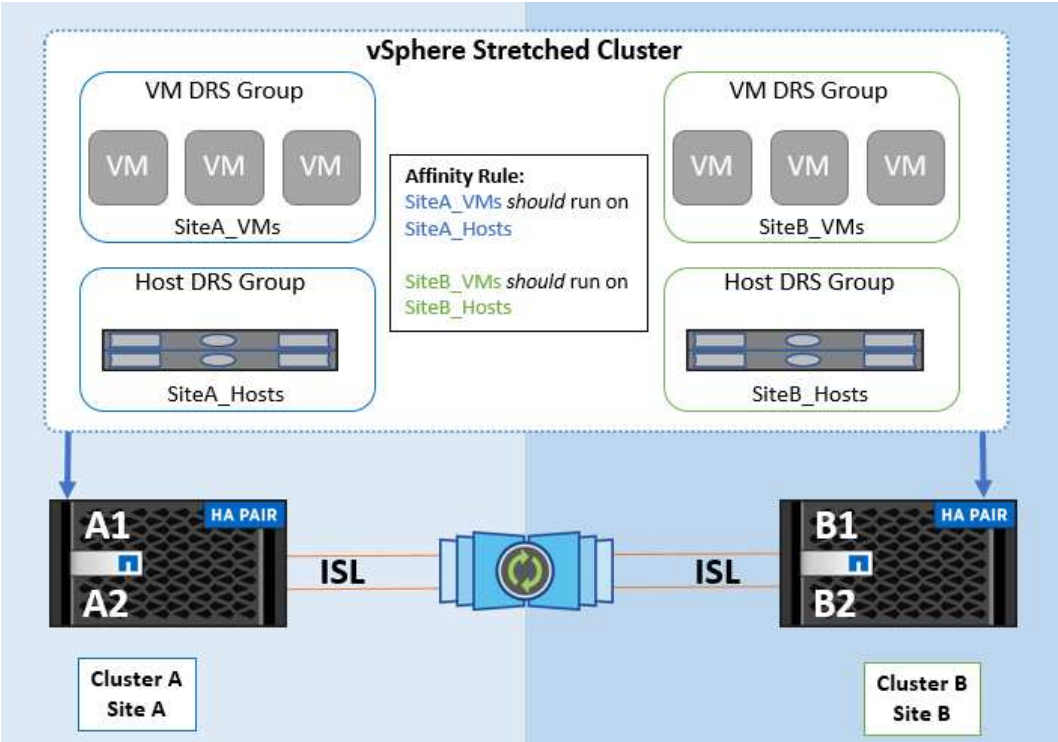
```
 To avoid cross-site traffic, NetApp recommends configuring DRS affinity
rules to manage a logical separation of VMs. This will ensure that unless
there is a complete site failure, HA and DRS will only use local hosts.
Si vous créez une règle d'affinité DRS pour votre cluster, vous pouvez
spécifier comment vSphere applique cette règle lors du basculement d'une
machine virtuelle.
```

Vous pouvez spécifier deux types de règles pour le basculement de vSphere HA :

- Les règles d'anti-affinité pour les machines virtuelles forcent les machines virtuelles spécifiées à rester séparées pendant les opérations de basculement.
- Les règles d'affinité des hôtes VM placent les machines virtuelles spécifiées sur un hôte particulier ou un membre d'un groupe défini d'hôtes lors des actions de basculement.

En utilisant les règles d'affinité pour les hôtes de machine virtuelle dans VMware DRS, il est possible d'avoir une séparation logique entre le site A et le site B, de sorte que la machine virtuelle s'exécute sur l'hôte au même site que la baie configurée comme contrôleur de lecture/écriture principal pour un datastore donné. De plus, les règles d'affinité des hôtes de VM permettent aux machines virtuelles de rester locales au stockage, ce qui à son tour ascert la connexion de la machine virtuelle en cas de défaillances réseau entre les sites.

Voici un exemple de groupes d'hôtes de machine virtuelle et de règles d'affinité.



Meilleure pratique

NetApp recommande de mettre en place des règles « à respecter » plutôt que des règles « à respecter », car elles sont violées par vSphere HA en cas de défaillance. L'utilisation de règles « must » peut entraîner des interruptions de service.

La disponibilité des services doit toujours prévaloir sur les performances. Dans le cas où un data Center

complet tombe en panne, les règles « must » doivent choisir des hôtes dans le groupe d'affinité des hôtes de la machine virtuelle, et lorsque le data Center n'est pas disponible, les machines virtuelles ne redémarrent pas.

Implémentation de VMware Storage DRS avec NetApp MetroCluster

La fonctionnalité VMware Storage DRS permet l'agrégation de datastores dans une seule unité et équilibre les disques de machines virtuelles lorsque les seuils de contrôle des E/S du stockage (SIOC) sont dépassés.

Le contrôle des E/S du stockage est activé par défaut sur les clusters DRS compatibles avec Storage DRS. Le contrôle des E/S du stockage permet à un administrateur de contrôler la quantité d'E/S de stockage allouée aux serveurs virtuels pendant les périodes d'encombrement des E/S. Ainsi, les serveurs virtuels plus importants sont préfénables aux serveurs virtuels moins importants pour l'allocation des ressources d'E/S.

Storage DRS utilise Storage vMotion pour migrer les machines virtuelles vers différents datastores au sein d'un cluster de datastores. Dans un environnement NetApp MetroCluster, la migration des machines virtuelles doit être contrôlée dans les datastores de ce site. Par exemple, la machine virtuelle A, qui s'exécute sur un hôte du site A, doit idéalement migrer au sein des datastores du SVM sur le site A. Si ce n'est pas le cas, la machine virtuelle continue à fonctionner mais avec des performances dégradées, puisque la lecture/l'écriture du disque virtuel se fera à partir du site B via des liens inter-sites.

*Lors de l'utilisation du stockage ONTAP, il est recommandé de désactiver Storage DRS.

- Storage DRS n'est généralement pas nécessaire ou recommandé pour une utilisation avec les systèmes de stockage ONTAP.
- ONTAP offre ses propres fonctionnalités d'efficacité du stockage, telles que la déduplication, la compression et la compaction, qui peuvent être affectées par Storage DRS.
- Si vous utilisez des snapshots ONTAP, Storage vMotion laisse derrière lui la copie de la machine virtuelle dans le snapshot, ce qui augmente potentiellement l'utilisation du stockage et peut avoir un impact sur les applications de sauvegarde telles que NetApp SnapCenter qui suivent les machines virtuelles et leurs snapshots ONTAP.

Directives de conception et de mise en œuvre VMSC

Ce document présente les lignes directrices en matière de conception et d'implémentation pour vMSC avec systèmes de stockage ONTAP.

Configuration du stockage NetApp

Les instructions de configuration pour NetApp MetroCluster sont disponibles à l'adresse "Documentation MetroCluster". Les instructions relatives à la synchronisation active SnapMirror (SMAs) sont également disponibles à l'adresse "Présentation de la continuité de l'activité SnapMirror".

Une fois que vous avez configuré MetroCluster, son administration revient à gérer un environnement ONTAP traditionnel. Vous pouvez configurer des machines virtuelles de stockage (SVM) à l'aide de divers outils tels que l'interface de ligne de commande (CLI), System Manager ou Ansible. Une fois les SVM configurés, créez des interfaces logiques (LIF), des volumes et des LUN sur le cluster qui seront utilisés pour les opérations normales. Ces objets seront automatiquement répliqués sur l'autre cluster à l'aide du réseau de peering de cluster.

Si vous n'utilisez pas MetroCluster ou si vous disposez de systèmes ONTAP qui ne sont pas pris en charge par MetroCluster, tels que les systèmes ASA r2, vous pouvez utiliser la synchronisation active SnapMirror qui offre une protection granulaire du datastore et un accès actif-actif sur plusieurs clusters ONTAP dans différents domaines de défaillance. Les directeurs de groupe utilisent des groupes de cohérence (CGS) pour assurer la

cohérence de l'ordre d'écriture dans un ou plusieurs datastores. Vous pouvez également créer plusieurs groupes de cohérence selon les besoins de votre application et de votre datastore. Les groupes de cohérence sont particulièrement utiles pour les applications qui nécessitent une synchronisation des données entre plusieurs datastores. Par exemple, des LVM invités sont distribués entre les datastores. Les SMAs prennent également en charge les mappages de périphériques Raw Device (RDM) et le stockage connecté par l'invité avec les initiateurs iSCSI invités. Pour en savoir plus sur les groupes de cohérence, rendez-vous sur "Présentation des groupes de cohérence".

La gestion d'une configuration vMSC avec SnapMirror Active Sync est différente de celle d'un MetroCluster. Tout d'abord, les SMAs sont une configuration SAN uniquement, aucun datastore NFS ne peut être protégé avec la synchronisation active SnapMirror. Ensuite, vous devez mapper les deux copies des LUN sur vos hôtes ESXi afin qu'elles puissent accéder aux datastores répliqués dans les deux domaines de défaillance. Troisièmement, vous devez créer un ou plusieurs groupes de cohérence pour les datastores à protéger avec la synchronisation active SnapMirror. Enfin, vous devez créer une politique de SnapMirror pour les groupes de cohérence que vous avez créés. Tout cela peut être facilement effectué à l'aide de l'assistant de protection du cluster du plug-in vCenter des outils ONTAP, ou manuellement via l'interface de ligne de commande ONTAP ou System Manager.

Utilisation du plug-in vCenter des outils ONTAP pour la synchronisation active SnapMirror

Le plug-in vCenter des outils ONTAP offre un moyen simple et intuitif de configurer la synchronisation active SnapMirror pour vMSC. Vous pouvez utiliser le plug-in vCenter des outils ONTAP pour créer et gérer des relations de synchronisation active SnapMirror entre deux clusters ONTAP. Ce plug-in fournit une interface facile à utiliser pour établir et gérer efficacement ces relations. Pour en savoir plus sur le plug-in vCenter des outils ONTAP, rendez-vous sur "Les outils ONTAP pour VMware vSphere" ou accédez directement à "Protégez à l'aide de la protection de cluster hôte".

Configuration de VMware vSphere

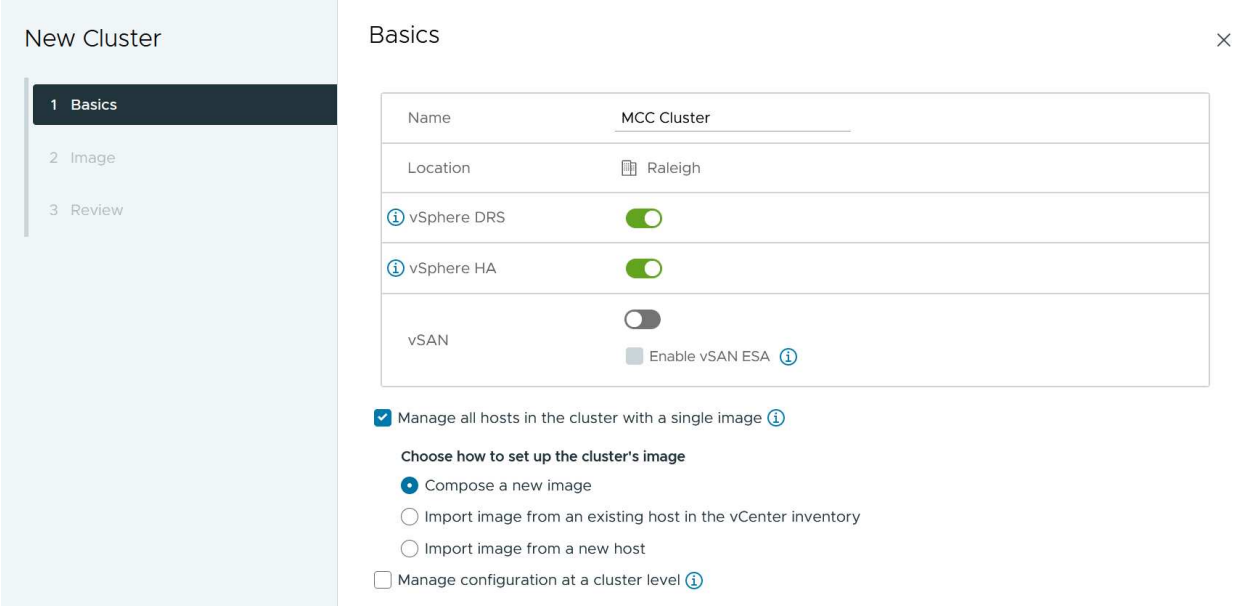
Créer un cluster haute disponibilité vSphere

La création d'un cluster vSphere HA est un processus en plusieurs étapes entièrement documenté à l'adresse "Comment créer et configurer des clusters dans vSphere client sur docs.vmware.com". En bref, vous devez d'abord créer un cluster vide, puis, à l'aide de vCenter, vous devez ajouter des hôtes et spécifier les paramètres vSphere HA et autres du cluster.

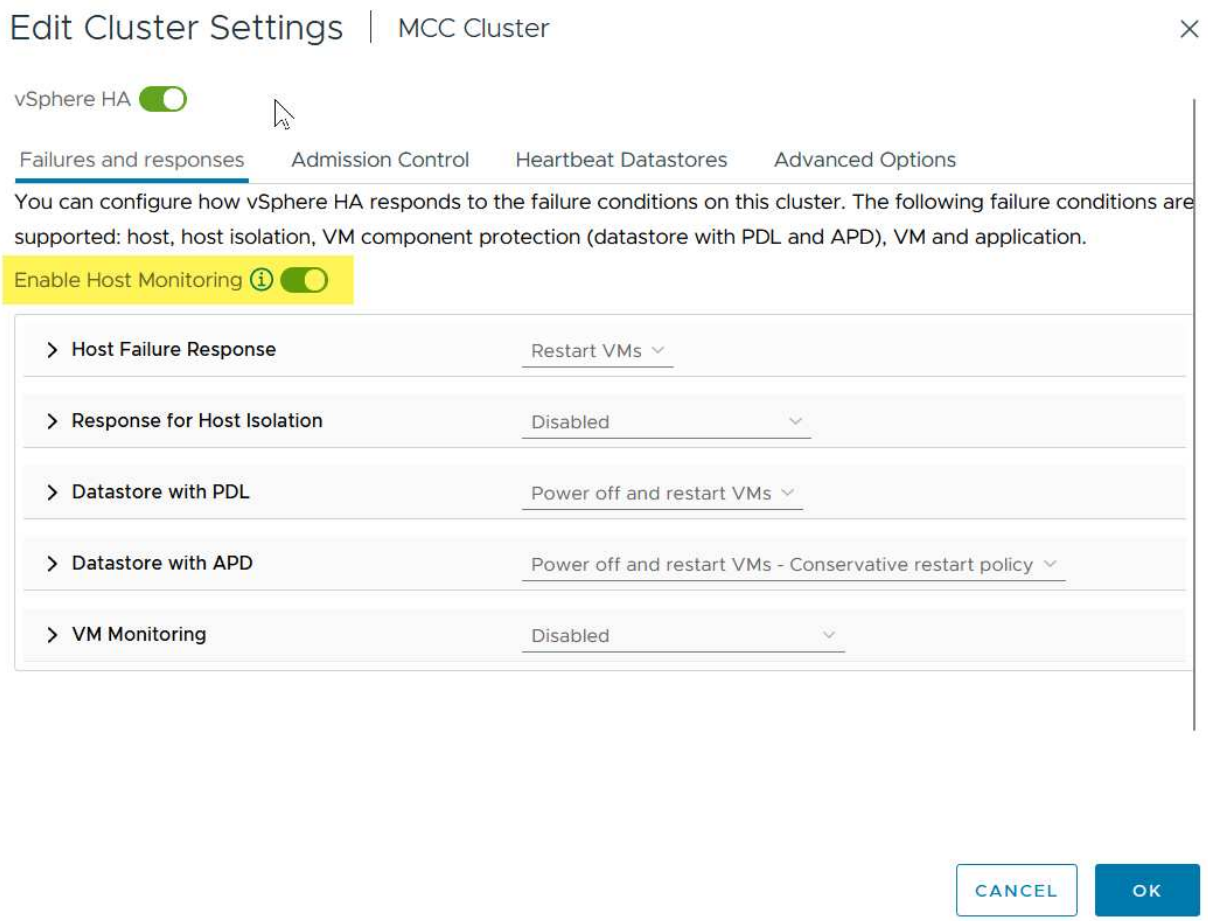
> Rien dans ce document ne remplace "Bonnes pratiques pour VMware vSphere Metro Storage Cluster". Ce contenu est fourni pour faciliter les références et ne remplace pas la documentation officielle de VMware.

Pour configurer un cluster HA, effectuez les étapes suivantes :

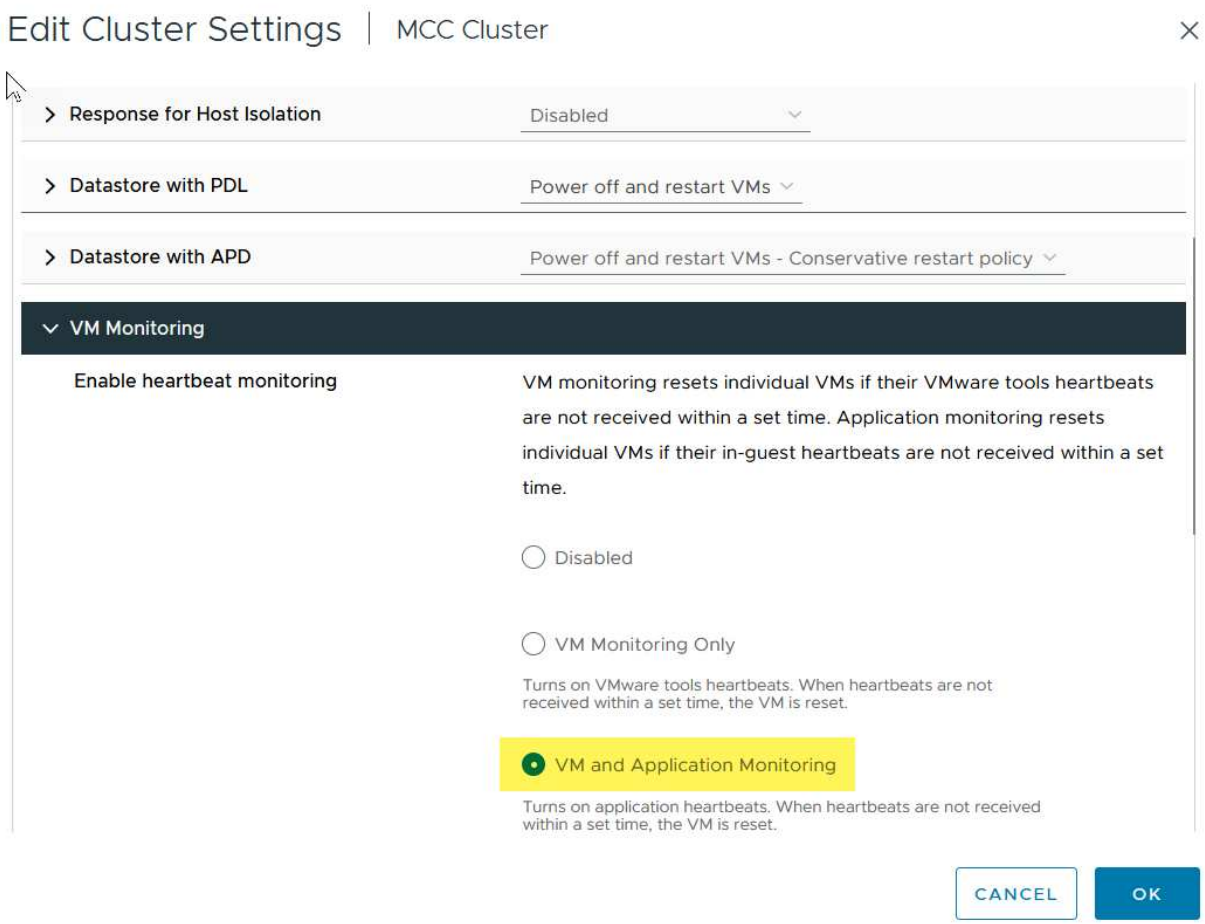
1. Connectez-vous à l'interface utilisateur vCenter.
2. Dans hôtes et clusters, accédez au data Center où vous souhaitez créer votre cluster haute disponibilité.
3. Cliquez avec le bouton droit de la souris sur l'objet de data Center et sélectionnez Nouveau cluster. Dans les notions de base, assurez-vous d'avoir activé vSphere DRS et vSphere HA. Suivez l'assistant.



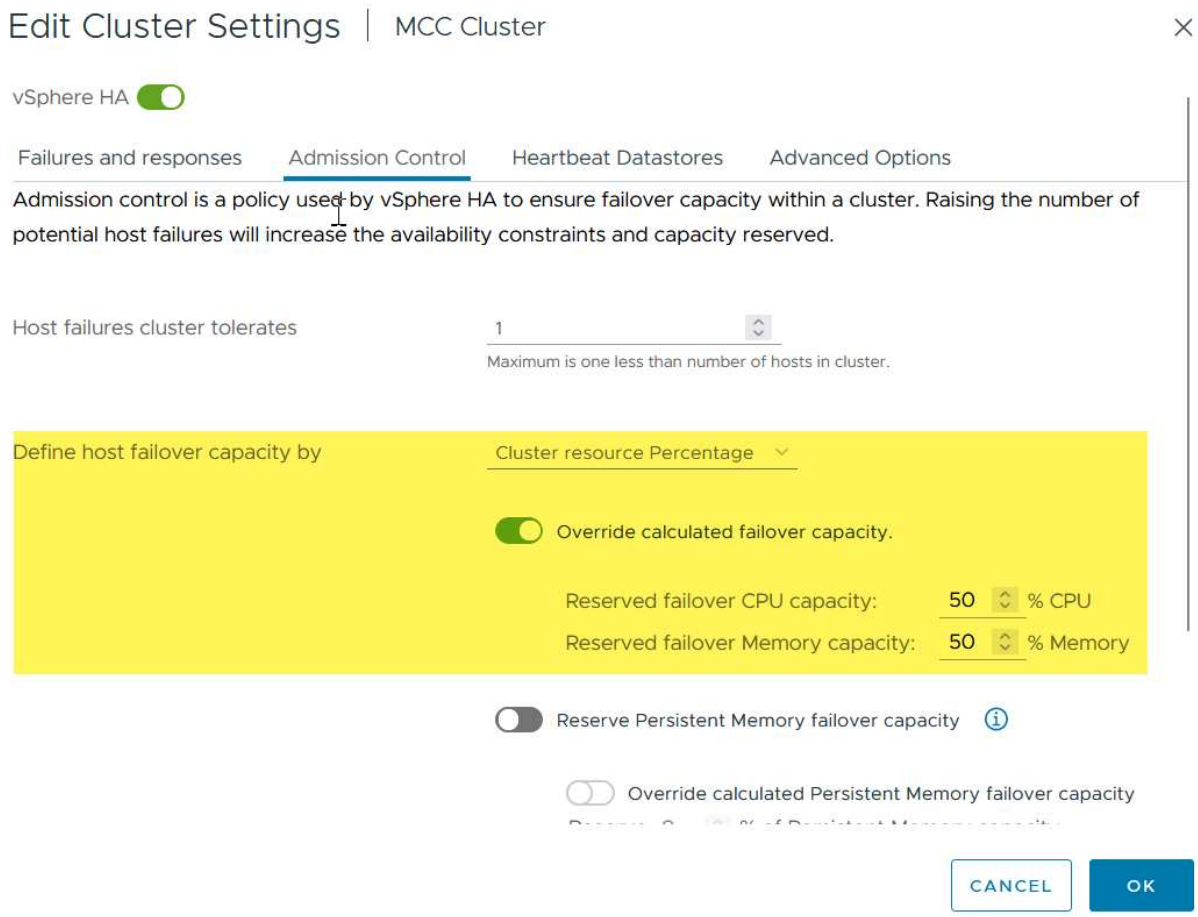
1. Sélectionnez le cluster et accédez à l'onglet configure. Sélectionnez vSphere HA et cliquez sur Edit.
2. Sous surveillance de l'hôte, sélectionnez l'option Activer la surveillance de l'hôte.



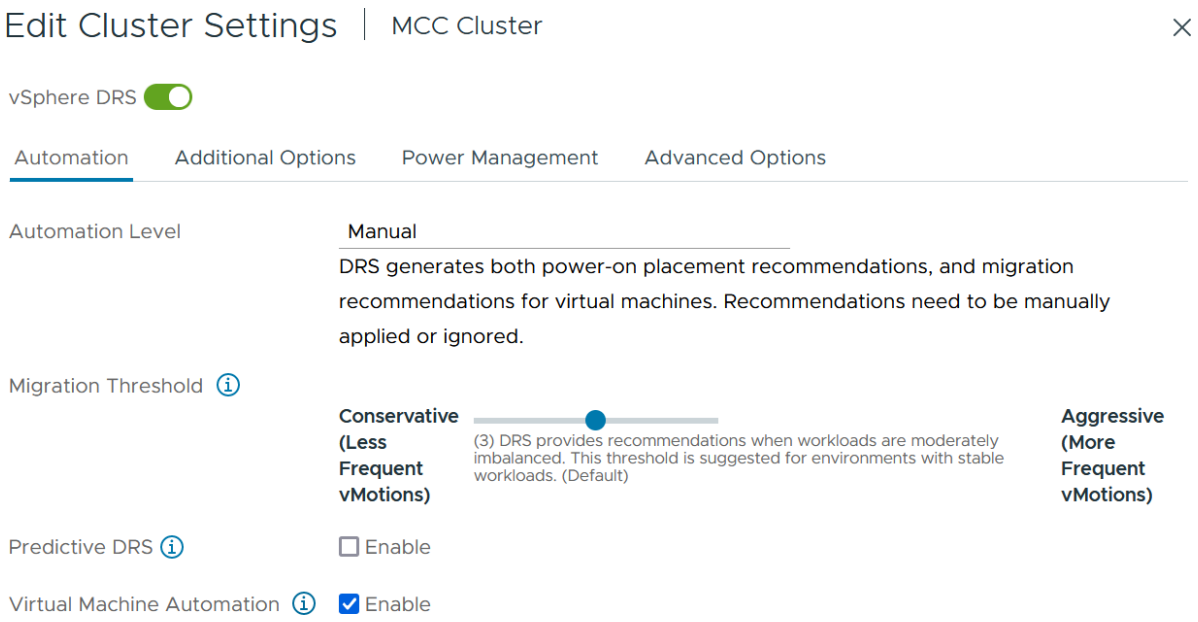
1. Toujours sous l'onglet défaillances et réponses, sous surveillance VM, sélectionnez l'option VM Monitoring Only ou VM and application Monitoring.



1. Sous contrôle d’admission, définissez l’option de contrôle d’admission HA sur réserve de ressources de cluster ; utilisez 50 % CPU/MEM.



1. Cliquez sur « OK ».
2. Sélectionnez DRS et cliquez sur EDIT.
3. Définissez le niveau d'automatisation sur manuel, sauf si vos applications en ont besoin.



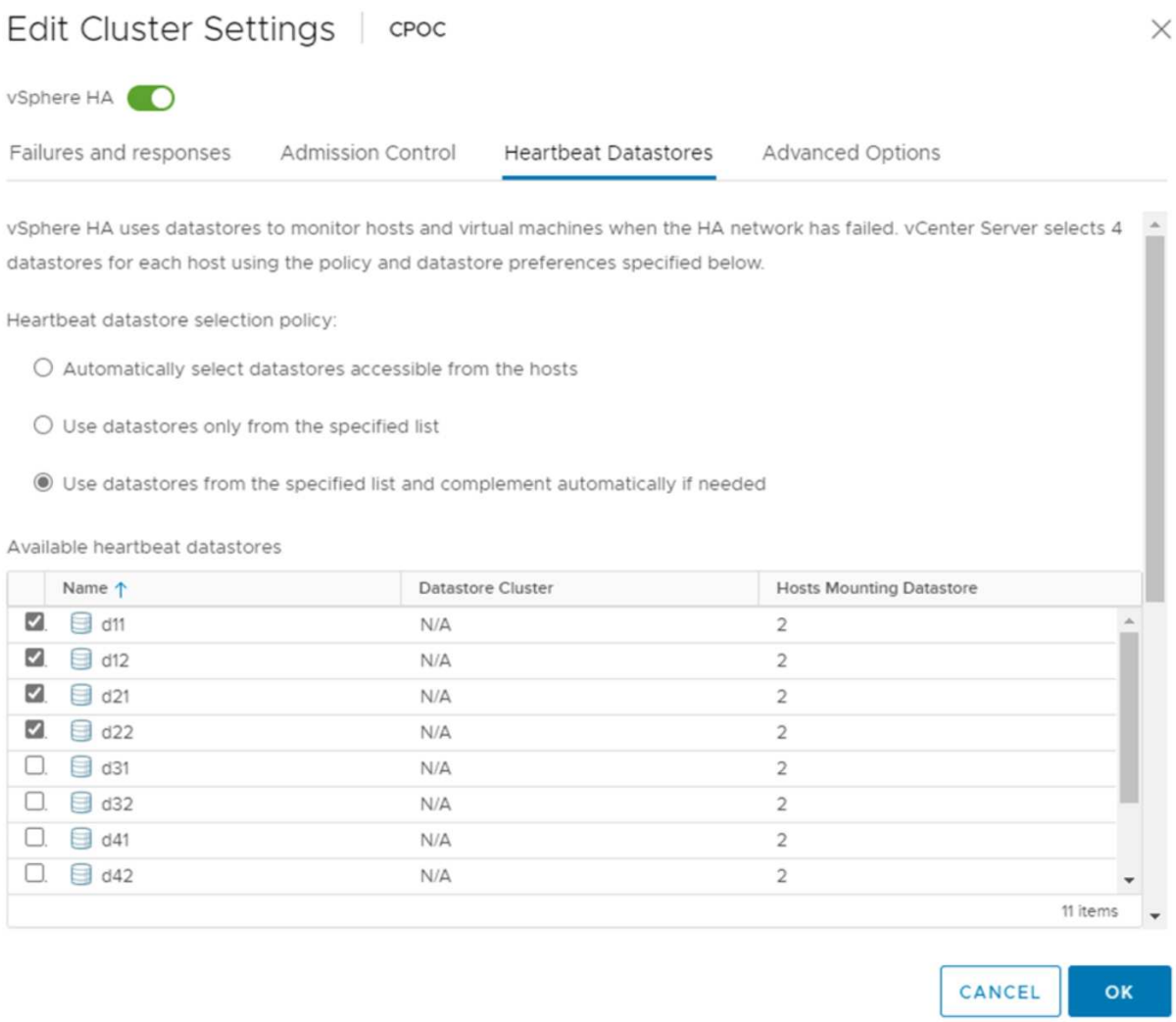
1. Activer la protection des composants VM, voir "docs.vmware.com".
2. Il est recommandé d'utiliser les paramètres vSphere HA supplémentaires suivants pour vMSC avec MetroCluster :

Panne	Réponse
Défaillance d'hôte	Redémarrage des machines virtuelles
Isolation de l'hôte	Désactivé
Datastore avec perte de périphérique permanente (PDL)	Mettez les machines virtuelles hors tension et redémarrez-les
Datastore avec tous les chemins en panne (APD)	Mettez les machines virtuelles hors tension et redémarrez-les
Client qui ne bat pas	Réinitialiser les VM
Règle de redémarrage de machine virtuelle	Déterminé par l'importance de la machine virtuelle
Réponse pour l'isolation de l'hôte	Arrêtez et redémarrez les machines virtuelles
Réponse pour datastore avec PDL	Mettez les machines virtuelles hors tension et redémarrez-les
Réponse pour le datastore avec APD	Mise hors tension et redémarrage des machines virtuelles (prudent)
Délai de basculement de machine virtuelle pour APD	3 minutes
Réponse pour la restauration APD avec délai d'expiration APD	Désactivé
Sensibilité de surveillance des machines virtuelles	Présélection haute

Configurez les datastores pour Heartbeat

VSphere HA utilise les datastores pour surveiller les hôtes et les machines virtuelles en cas de panne du réseau de gestion. Vous pouvez configurer la façon dont vCenter sélectionne les datastores Heartbeat. Pour configurer des datastores pour les pulsations, procédez comme suit :

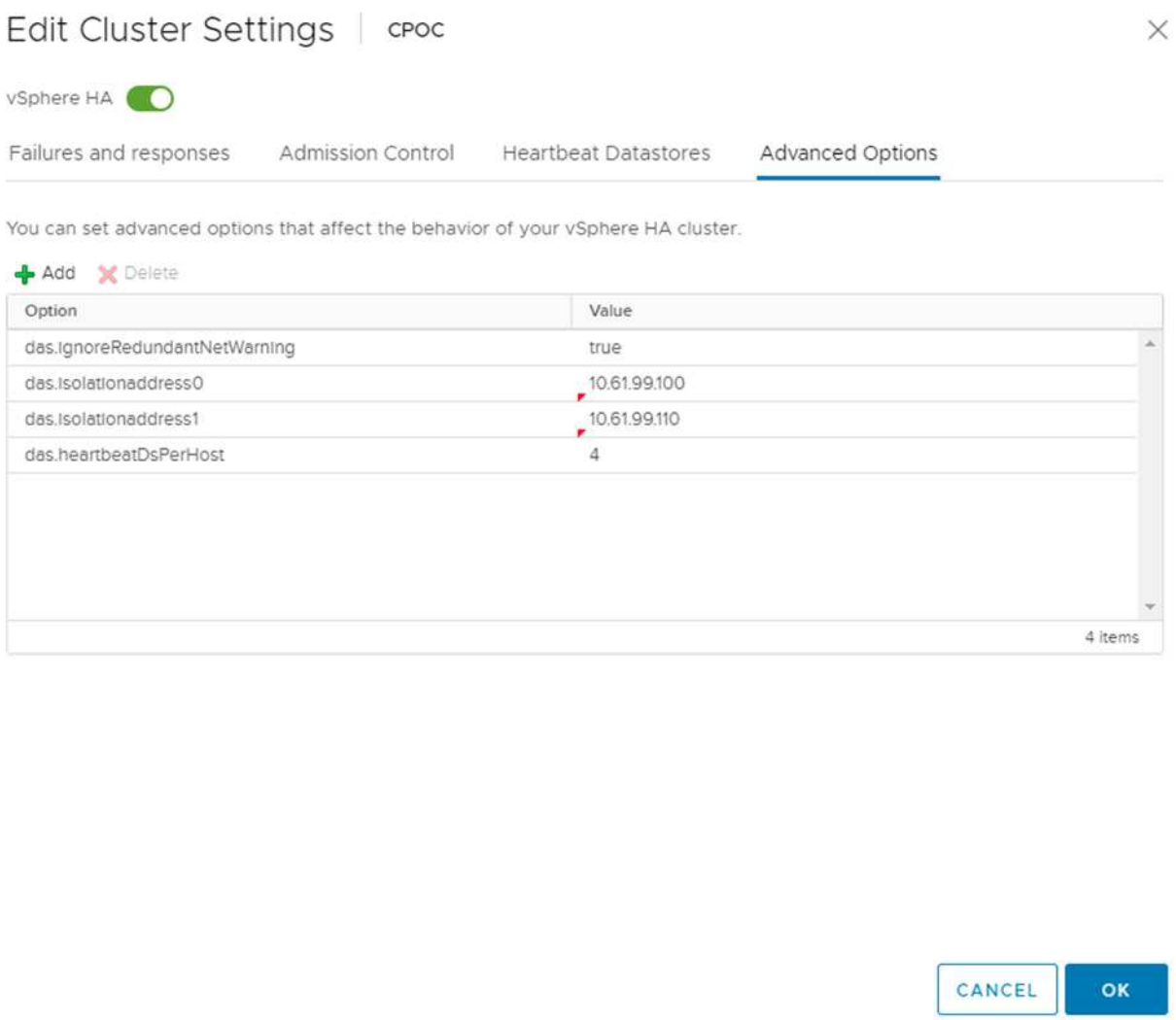
1. Dans la section pulsation du datastore, sélectionnez utiliser les datastores dans la liste spécifiée et complétez automatiquement si nécessaire.
2. Sélectionnez les datastores que vCenter doit utiliser sur les deux sites et appuyez sur OK.



Configurer les options avancées

Les événements d'isolation se produisent lorsque les hôtes d'un cluster haute disponibilité perdent la connectivité au réseau ou à d'autres hôtes du cluster. Par défaut, vSphere HA utilise la passerelle par défaut de son réseau de gestion comme adresse d'isolation par défaut. Toutefois, vous pouvez spécifier des adresses d'isolement supplémentaires pour que l'hôte puisse envoyer une requête ping afin de déterminer si une réponse d'isolement doit être déclenchée. Ajoutez deux adresses IP d'isolation pouvant être ping, une par site. N'utilisez pas l'adresse IP de la passerelle. Le paramètre avancé de vSphere HA utilisé est das.isolaaddress. Vous pouvez utiliser des adresses IP ONTAP ou Mediator à cette fin.

Pour plus d'informations, reportez-vous à la section "Bonnes pratiques pour VMware vSphere Metro Storage Cluster"._



L'ajout d'un paramètre avancé appelé das.heartbeatDsPerHost peut augmenter le nombre de datastores de pulsation. Utilisez quatre datastores de pulsation (DSS HB)—deux par site. Utilisez l'option « Sélectionner dans la liste mais compléter ». Ceci est nécessaire car si un site tombe en panne, vous avez toujours besoin de deux DSS HB. Toutefois, celles-ci n'ont pas à être protégées avec la synchronisation active MetroCluster ou SnapMirror.

Pour plus d'informations, reportez-vous à la section "Bonnes pratiques pour VMware vSphere Metro Storage Cluster"._

Affinité avec VMware DRS pour NetApp MetroCluster

Dans cette section, nous créons des groupes DRS pour les machines virtuelles et les hôtes pour chaque site/cluster dans l'environnement MetroCluster. Ensuite, nous configurons les règles VM/Host pour aligner l'affinité des hôtes VM avec les ressources de stockage locales. Par exemple, les machines virtuelles du site A appartiennent au groupe de machines virtuelles sitea_VM et les hôtes du site A appartiennent au groupe d'hôtes sitea_hosts. Ensuite, dans VM\Host Rules, nous faisons état que sitea_vm doit s'exécuter sur les hôtes de sitea_hosts.

- NetApp recommande vivement la spécification **devrait s'exécuter sur les hôtes du groupe** plutôt que la spécification **doit s'exécuter sur les hôtes du groupe**. En cas de défaillance d'un hôte sur un site, les machines virtuelles Du site A doivent être redémarrées sur les hôtes du site B via vSphere HA, mais cette dernière spécification ne permet pas à HA de redémarrer les machines virtuelles sur le site B, car il s'agit d'une règle stricte. Il s'agit d'une règle souple qui ne sera pas respectée en cas de haute disponibilité, garantissant ainsi la disponibilité plutôt que la performance.
- Vous pouvez créer une alarme basée sur des événements qui est déclenchée lorsqu'une machine virtuelle viole une règle d'affinité VM-Host. Dans le client vSphere, ajoutez une nouvelle alarme pour la machine virtuelle et sélectionnez « VM viole VM-Host Affinity Rule » comme déclencheur d'événement. Pour plus d'informations sur la création et la modification d'alarmes, reportez-vous à la "Surveillance et performances vSphere" documentation.

Créer des groupes d'hôtes DRS

Pour créer des groupes d'hôtes DRS spécifiques au site A et au site B, procédez comme suit :

1. Dans le client Web vSphere, cliquez avec le bouton droit de la souris sur le cluster dans l'inventaire et sélectionnez Paramètres.
2. Cliquez sur VM\Host Groups.
3. Cliquez sur Ajouter.
4. Saisissez le nom du groupe (par exemple, sitea_hosts).
5. Dans le menu Type, sélectionnez Groupe d'hôtes.
6. Cliquez sur Ajouter et sélectionnez les hôtes souhaités sur le site A, puis cliquez sur OK.
7. Répétez ces étapes pour ajouter un autre groupe d'hôtes pour le site B.
8. Cliquez sur OK.

Créer des groupes VM DRS

Pour créer des groupes VM DRS spécifiques au site A et au site B, procédez comme suit :

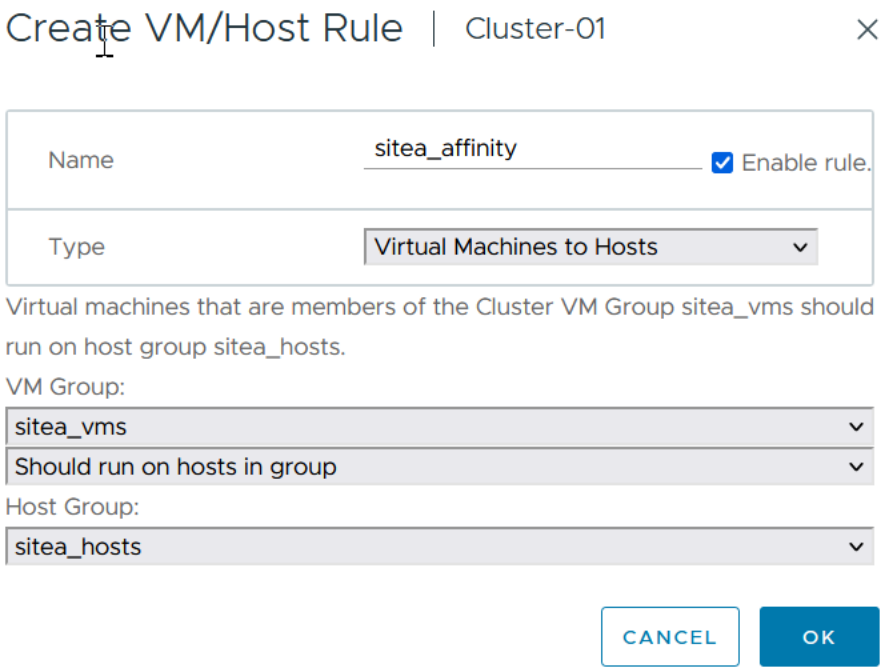
1. Dans le client Web vSphere, cliquez avec le bouton droit de la souris sur le cluster dans l'inventaire et sélectionnez Paramètres.
2. Cliquez sur VM\Host Groups.
3. Cliquez sur Ajouter.
4. Saisissez le nom du groupe (par exemple, sitea_vm).
5. Dans le menu Type, sélectionnez VM Group.
6. Cliquez sur Ajouter, sélectionnez les machines virtuelles souhaitées sur le site A, puis cliquez sur OK.
7. Répétez ces étapes pour ajouter un autre groupe d'hôtes pour le site B.
8. Cliquez sur OK.

Créer des règles d'hôte VM

Pour créer des règles d'affinité DRS spécifiques au site A et au site B, procédez comme suit :

1. Dans le client Web vSphere, cliquez avec le bouton droit de la souris sur le cluster dans l'inventaire et sélectionnez Paramètres.

2. Cliquez sur VM\Host Rules.
3. Cliquez sur Ajouter.
4. Tapez le nom de la règle (par exemple, sitea_affinité).
5. Vérifiez que l'option Activer la règle est cochée.
6. Dans le menu Type, sélectionnez ordinateurs virtuels vers hôtes.
7. Sélectionnez le groupe VM (par exemple, sitea_vm).
8. Sélectionnez le groupe Host (par exemple, sitea_hosts).
9. Répétez ces étapes pour ajouter une autre règle VM\Host pour le site B.
10. Cliquez sur OK.



Créez des clusters de datastores si nécessaire

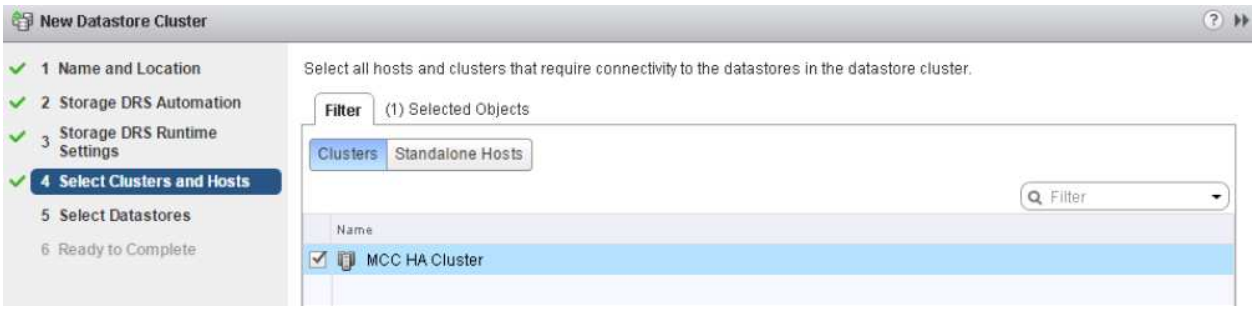
Pour configurer un cluster de datastore pour chaque site, procédez comme suit :

1. À l'aide du client web vSphere, accédez au data Center où réside le cluster HA sous Storage.
2. Cliquez avec le bouton droit de la souris sur l'objet datacenter et sélectionnez Storage > New datastore Cluster.

*Lors de l'utilisation du stockage ONTAP, il est recommandé de désactiver Storage DRS.

- Storage DRS n'est généralement pas nécessaire ou recommandé pour une utilisation avec les systèmes de stockage ONTAP.
- ONTAP offre ses propres fonctionnalités d'efficacité du stockage, telles que la déduplication, la compression et la compaction, qui peuvent être affectées par Storage DRS.
- Si vous utilisez des snapshots ONTAP, Storage vMotion laisse derrière lui la copie de la machine virtuelle dans le snapshot, ce qui augmente potentiellement l'utilisation du stockage et peut avoir un impact sur les applications de sauvegarde telles que NetApp SnapCenter qui suivent les machines virtuelles et leurs snapshots ONTAP.

1. Sélectionnez le cluster HA et cliquez sur Next.



1. Sélectionnez les datastores appartenant au site A et cliquez sur Suivant.

1. Vérifiez les options et cliquez sur Terminer.
2. Répétez ces étapes pour créer le cluster de datastore du site B et vérifier que seuls les datastores du site B sont sélectionnés.

Disponibilité du serveur vCenter

Vos appliances vCenter Server (VCSA) doivent être protégées avec vCenter HA. VCenter HA vous permet de déployer deux VCSA dans une paire haute disponibilité actif-passif. Un dans chaque domaine de défaillance. Pour en savoir plus sur vCenter HA, rendez-vous sur "docs.vmware.com".

Résilience pour les événements planifiés et non planifiés

NetApp MetroCluster et la synchronisation active SnapMirror sont des outils puissants qui améliorent la haute disponibilité et la continuité de l'activité du matériel NetApp et du logiciel ONTAP®.

Ces outils assurent une protection à l'échelle du site pour l'ensemble de l'environnement de stockage, garantissant ainsi la disponibilité permanente de vos données. Que vous utilisiez des serveurs autonomes, des clusters de serveurs à haute disponibilité, des conteneurs ou des serveurs virtualisés, la technologie NetApp assure la disponibilité du stockage de manière transparente en cas de panne totale due à une coupure d'alimentation, à des problèmes de climatisation, de connectivité réseau, à l'arrêt des baies de stockage ou à une erreur opérationnelle.

La synchronisation active MetroCluster et SnapMirror propose trois méthodes de base pour la continuité des données en cas d'événements planifiés ou non :

- Des composants redondants pour une protection contre les défaillances d'un seul composant
- Basculement de haute disponibilité locale en cas d'événements affectant un contrôleur unique
- Protection complète du site – reprise rapide du service en déplaçant le stockage et l'accès client du cluster source vers le cluster de destination

Cela signifie que les opérations se poursuivent en toute transparence en cas de défaillance d'un seul composant et reviennent automatiquement au fonctionnement redondant lorsque le composant défectueux est remplacé.

Tous les clusters ONTAP, à l'exception des clusters à un seul nœud (en général, les versions Software-defined, telles que ONTAP Select, par exemple), disposent de fonctionnalités haute disponibilité intégrées appelées Takeover et giveback. Chaque contrôleur du cluster est couplé à un autre contrôleur, formant une paire haute disponibilité. Ces paires garantissent que chaque nœud est connecté localement au stockage.

Le basculement est un processus automatisé qui consiste à prendre le contrôle du stockage d'un nœud pour assurer les services de données. Le rétablissement est le processus inverse qui restaure le fonctionnement normal. Le basculement peut être planifié, par exemple lors de la maintenance matérielle ou des mises à niveau ONTAP, ou non planifié, suite à une panne matérielle ou de panique sur un nœud.

Lors d'un basculement, les LIF NAS dans les configurations MetroCluster basculent automatiquement. Toutefois, les LIFs SAN ne basculent pas ; elles continueront d'utiliser le chemin direct vers les LUN (Logical Unit Numbers).

Pour plus d'informations sur le basculement et le rétablissement HA, consultez le "Présentation de la gestion des paires HAUTE DISPONIBILITÉ". Notez que cette fonctionnalité n'est pas spécifique à la synchronisation active MetroCluster ou SnapMirror.

Le basculement de site avec MetroCluster a lieu lorsqu'un site est hors ligne ou lors d'une activité planifiée pour la maintenance à l'échelle du site. Le site restant assume la propriété des ressources de stockage (disques et agrégats) du cluster hors ligne, et les SVM sur le site en panne sont mis en ligne et redémarrés sur le site en cas de sinistre, tout en préservant leur identité complète pour l'accès des clients et des hôtes.

Avec la synchronisation active SnapMirror, dans la mesure où les deux copies sont activement utilisées simultanément, vos hôtes existants continueront de fonctionner. Le médiateur ONTAP est nécessaire pour garantir que le basculement de site se produit correctement.

Scénarios de défaillance pour vMSC avec MetroCluster

Les sections suivantes décrivent les résultats attendus de différents scénarios de défaillance avec les systèmes vMSC et NetApp MetroCluster.

Défaillance d’un seul chemin de stockage

Dans ce scénario, si des composants tels que le port HBA, le port réseau, le port du commutateur de données frontal ou un câble FC ou Ethernet échouent, ce chemin particulier vers le périphérique de stockage est marqué comme mort par l’hôte ESXi. Si plusieurs chemins sont configurés pour le périphérique de stockage en fournissant la résilience au niveau du port HBA/réseau/commutateur, ESXi effectue idéalement un basculement de chemin. Pendant cette période, les ordinateurs virtuels restent en fonctionnement sans être affectés, car la disponibilité du stockage est assurée par plusieurs chemins vers le périphérique de stockage.

> Dans ce scénario, le comportement de MetroCluster n’a pas changé et tous les datastores sont toujours intacts sur leurs sites respectifs.

Meilleure pratique

Dans les environnements dans lesquels les volumes NFS/iSCSI sont utilisés, NetApp recommande de configurer au moins deux liaisons montantes réseau pour le port vmkernel NFS dans le vSwitch standard et la même pour le groupe de ports où l’interface vmkernel NFS est mappée pour le vSwitch distribué. Le regroupement de cartes réseau peut être configuré en mode actif-actif ou actif-veille.

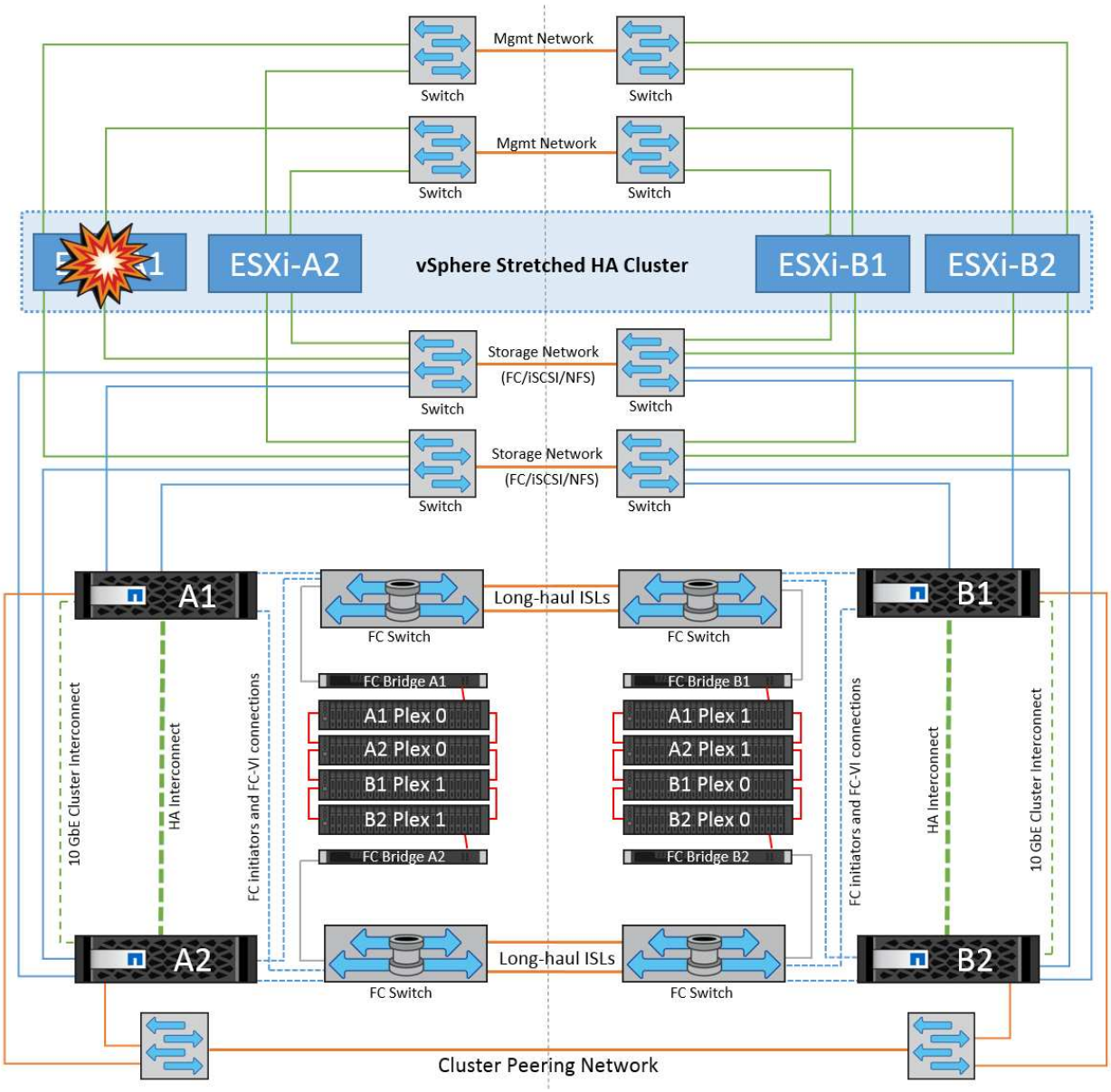
En outre, pour les LUN iSCSI, les chemins d’accès multiples doivent être configurés en liant les interfaces vmkernel aux adaptateurs réseau iSCSI. Pour plus d’informations, reportez-vous à la documentation sur le stockage vSphere.

Meilleure pratique

Dans les environnements dans lesquels des LUN Fibre Channel sont utilisées, NetApp recommande d’avoir au moins deux HBA, ce qui garantit la résilience au niveau des HBA/ports. NetApp recommande également la segmentation entre un initiateur unique et une seule cible comme meilleure pratique pour la configuration de la segmentation.

Virtual Storage Console (VSC) doit être utilisé pour définir des règles de chemins d’accès multiples, car il définit des règles pour tous les périphériques de stockage NetApp, nouveaux ou existants.

Défaillance d’un hôte ESXi unique



Dans ce scénario, en cas de défaillance de l'hôte ESXi, le nœud maître du cluster VMware HA détecte la panne de l'hôte, car il ne reçoit plus de pulsations réseau. Pour déterminer si l'hôte est réellement en panne ou uniquement une partition réseau, le nœud maître surveille les pulsations du datastore et, s'il est absent, il effectue une vérification finale en envoyant une requête ping aux adresses IP de gestion de l'hôte en panne. Si toutes ces vérifications sont négatives, le nœud maître déclare cet hôte comme étant en panne et toutes les machines virtuelles qui s'exécutaient sur cet hôte en panne sont redémarrées sur l'hôte survivant du cluster.

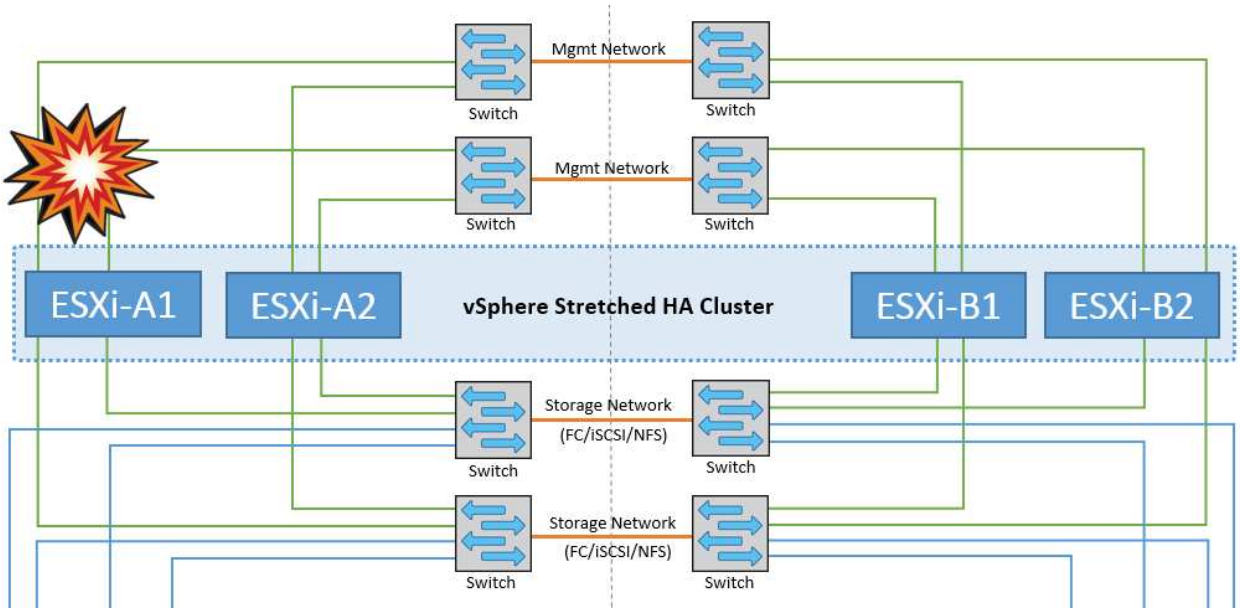
Si les règles d'affinité des machines virtuelles DRS et des hôtes ont été configurées (les machines virtuelles du groupe de machines virtuelles sitea_vm doivent exécuter des hôtes dans le groupe d'hôtes sitea_hosts), le maître haute disponibilité vérifie d'abord les ressources disponibles sur le site A. Si aucun hôte n'est disponible sur le site A, le maître tente de redémarrer les machines virtuelles sur les hôtes du site B.

Il est possible que les machines virtuelles soient démarrées sur les hôtes ESXi de l'autre site s'il existe une contrainte de ressource sur le site local. Cependant, les règles d'affinité VM et hôte DRS définies seront correctes si des règles sont enfreintes en migrant les machines virtuelles vers des hôtes ESXi survivants sur le site local. Dans les cas où DRS est défini sur manuel, NetApp recommande d'invoquer DRS et d'appliquer les recommandations pour corriger le positionnement de la machine virtuelle.

Dans ce scénario, le comportement de MetroCluster n'a pas changé et tous les datastores sont toujours

intacts sur leurs sites respectifs.

Isolation de l'hôte ESXi

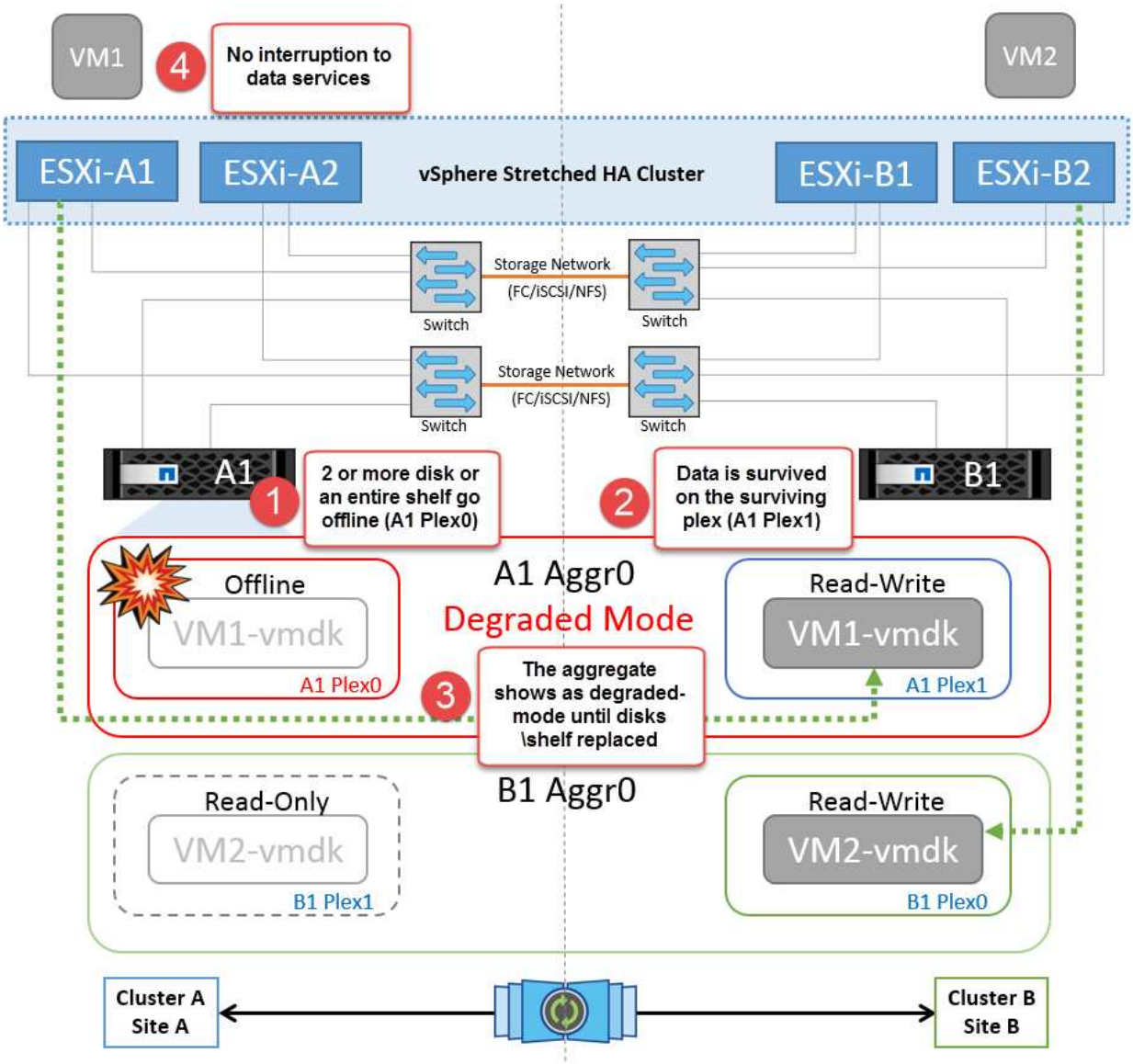


Dans ce scénario, si le réseau de gestion de l'hôte ESXi est en panne, le nœud principal du cluster HA ne recevra aucun battement de cœur. Cet hôte est donc isolé dans le réseau. Pour déterminer s'il a échoué ou s'il est isolé uniquement, le nœud maître commence à surveiller le battement de cœur du datastore. S'il est présent, l'hôte est déclaré isolé par le nœud maître. Selon la réponse d'isolement configurée, l'hôte peut choisir de mettre hors tension, d'arrêter les machines virtuelles ou même de laisser les machines virtuelles sous tension. L'intervalle par défaut pour la réponse d'isolement est de 30 secondes.

Dans ce scénario, le comportement de MetroCluster n'a pas changé et tous les datastores sont toujours intacts sur leurs sites respectifs.

Panne de tiroir disque

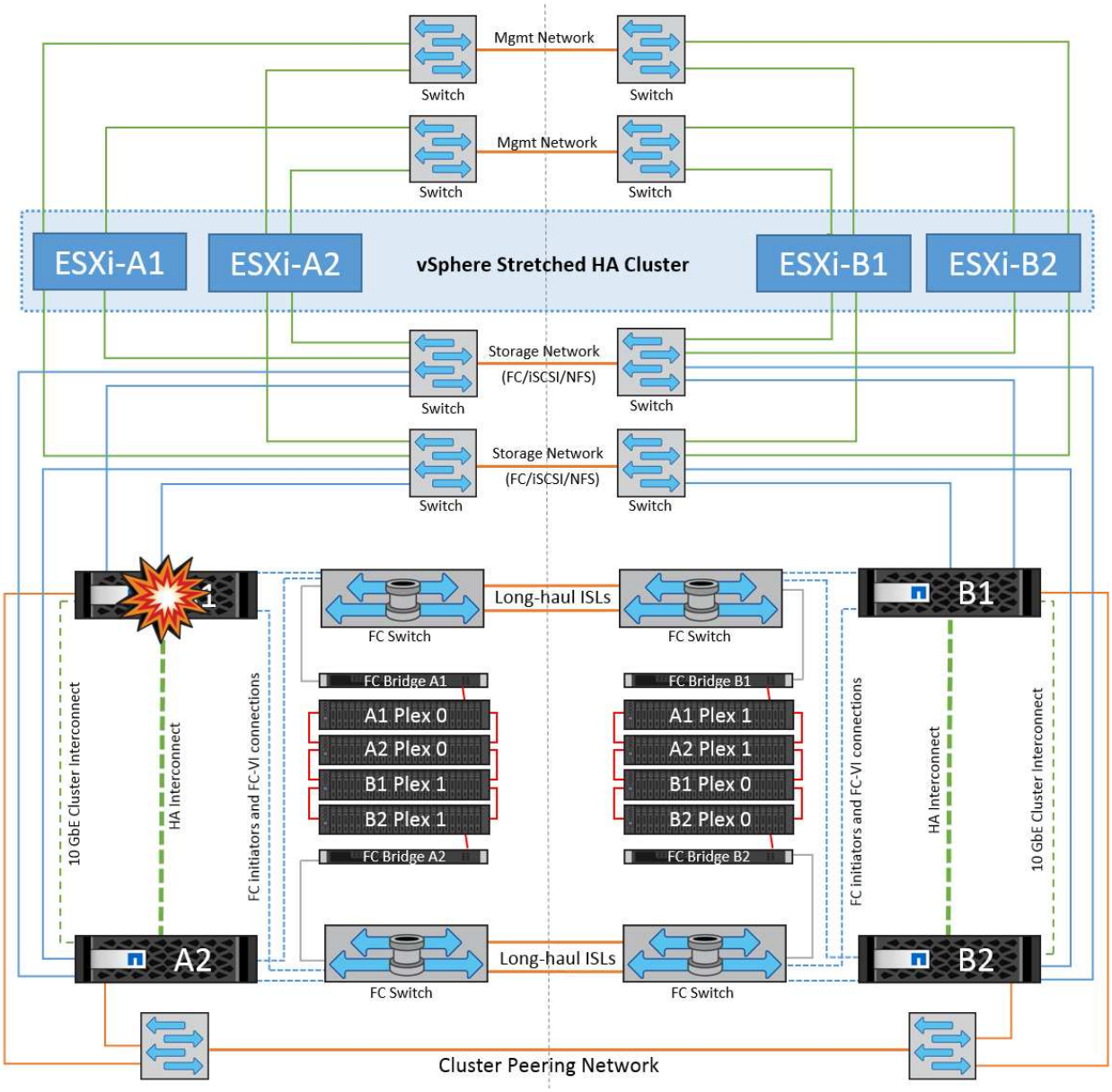
Dans ce scénario, il y a une panne de plus de deux disques ou d'un tiroir entier. Les données sont servies depuis le plex opérationnel sans interruption des services de données. La défaillance de disque peut affecter un plex local ou distant. Les agrégats s'affichent en mode dégradé, car un seul plex est actif. Une fois les disques défaillants remplacés, les agrégats affectés resynchroniseront automatiquement pour reconstruire les données. Après la resynchronisation, les agrégats reviennent automatiquement en mode miroir normal. Si plus de deux disques d'un même groupe RAID sont défectueux, le plex doit être reconstruit.



- [NOTE]
- Au cours de cette période, les opérations d'E/S du serveur virtuel ne sont pas affectées, mais les performances sont réduites du fait de l'accès aux données depuis le tiroir disque distant via des liens ISL.

Panne d'un seul contrôleur de stockage

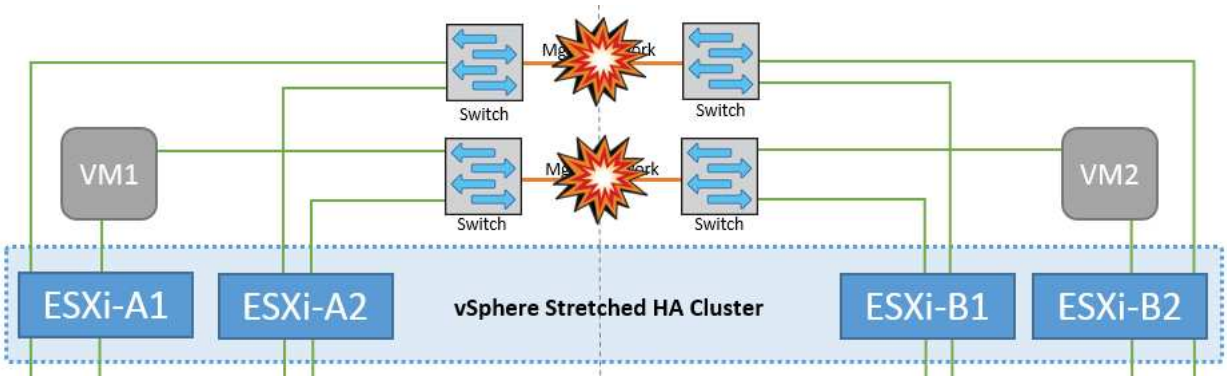
Dans ce scénario, l'un des deux contrôleurs de stockage tombe en panne sur un site. Comme il existe une paire haute disponibilité sur chaque site, la panne d'un nœud entraîne le basculement vers l'autre nœud de manière transparente et automatique. Par exemple, si le nœud A1 tombe en panne, son stockage et ses charges de travail sont automatiquement transférés vers le nœud A2. Les machines virtuelles ne seront pas affectées, car tous les plexes restent disponibles. Les nœuds du second site (B1 et B2) ne sont pas affectés. En outre, vSphere HA ne prendra aucune action, car le nœud maître du cluster recevra toujours les battements de cœur du réseau.



Si le basculement fait partie d'un incident en cours (le nœud A1 bascule vers A2) et qu'il y a une panne ultérieure de A2, ou la panne complète du site A, le basculement après un incident peut se produire sur le site B.

Défaillances de liaison entre commutateurs

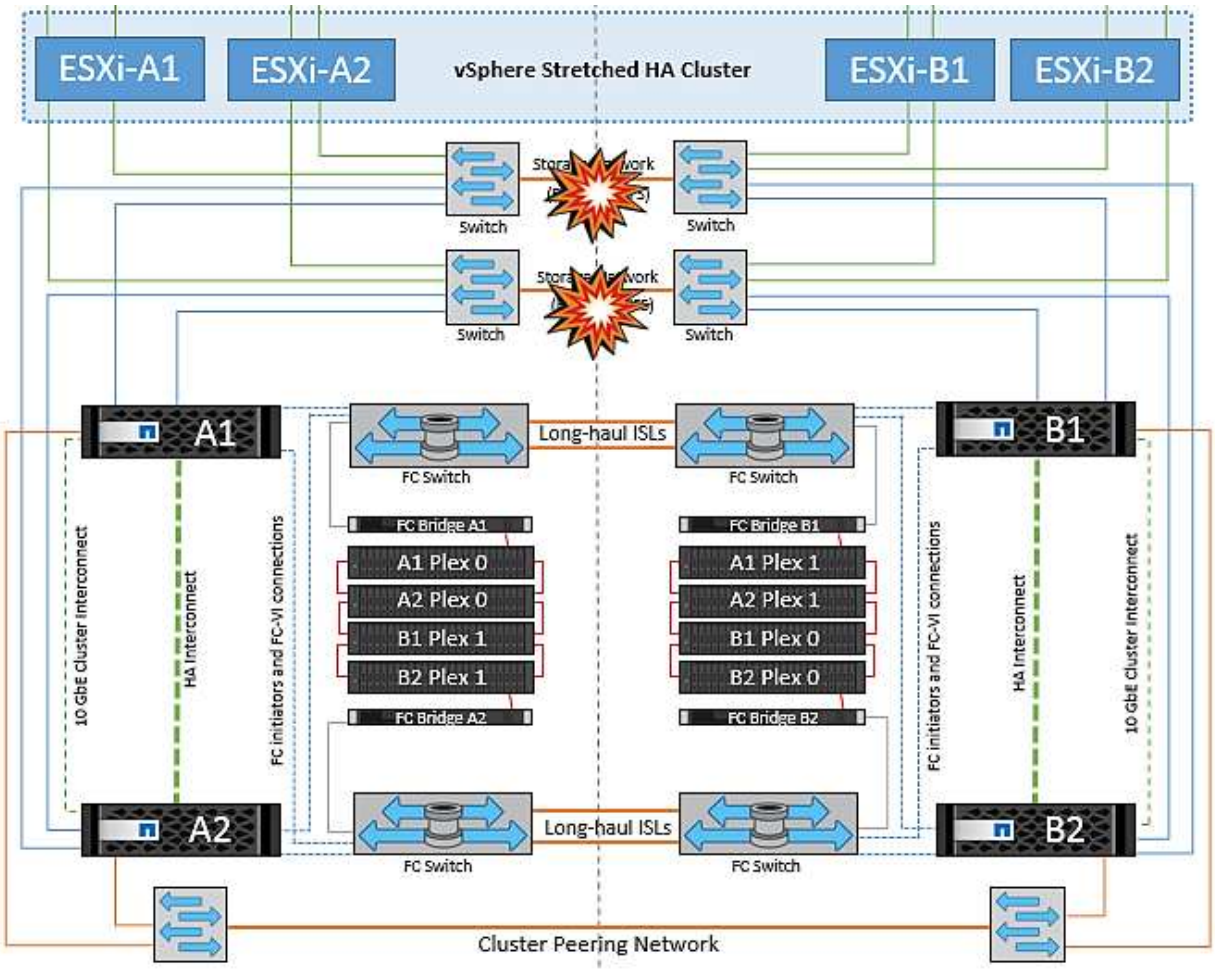
Défaillance de la liaison inter-commutateur sur le réseau de gestion



Dans ce scénario, si les liaisons ISL du réseau de gestion de l'hôte frontal tombent en panne, les hôtes ESXi du site A ne pourront pas communiquer avec les hôtes ESXi du site B. Cela entraîne une partition réseau, car les hôtes ESXi d'un site particulier ne peuvent pas envoyer les battements de cœur du réseau au nœud maître du cluster HA. Ainsi, il y aura deux segments de réseau en raison de la partition et il y aura un nœud maître dans chaque segment qui protégera les machines virtuelles des défaillances de l'hôte au sein du site particulier.

> Pendant cette période, les machines virtuelles restent en cours d'exécution et le comportement de MetroCluster n'a pas changé dans ce scénario. Tous les datastores sont toujours intacts sur leurs sites respectifs.

Défaillance de la liaison intercommutateur sur le réseau de stockage

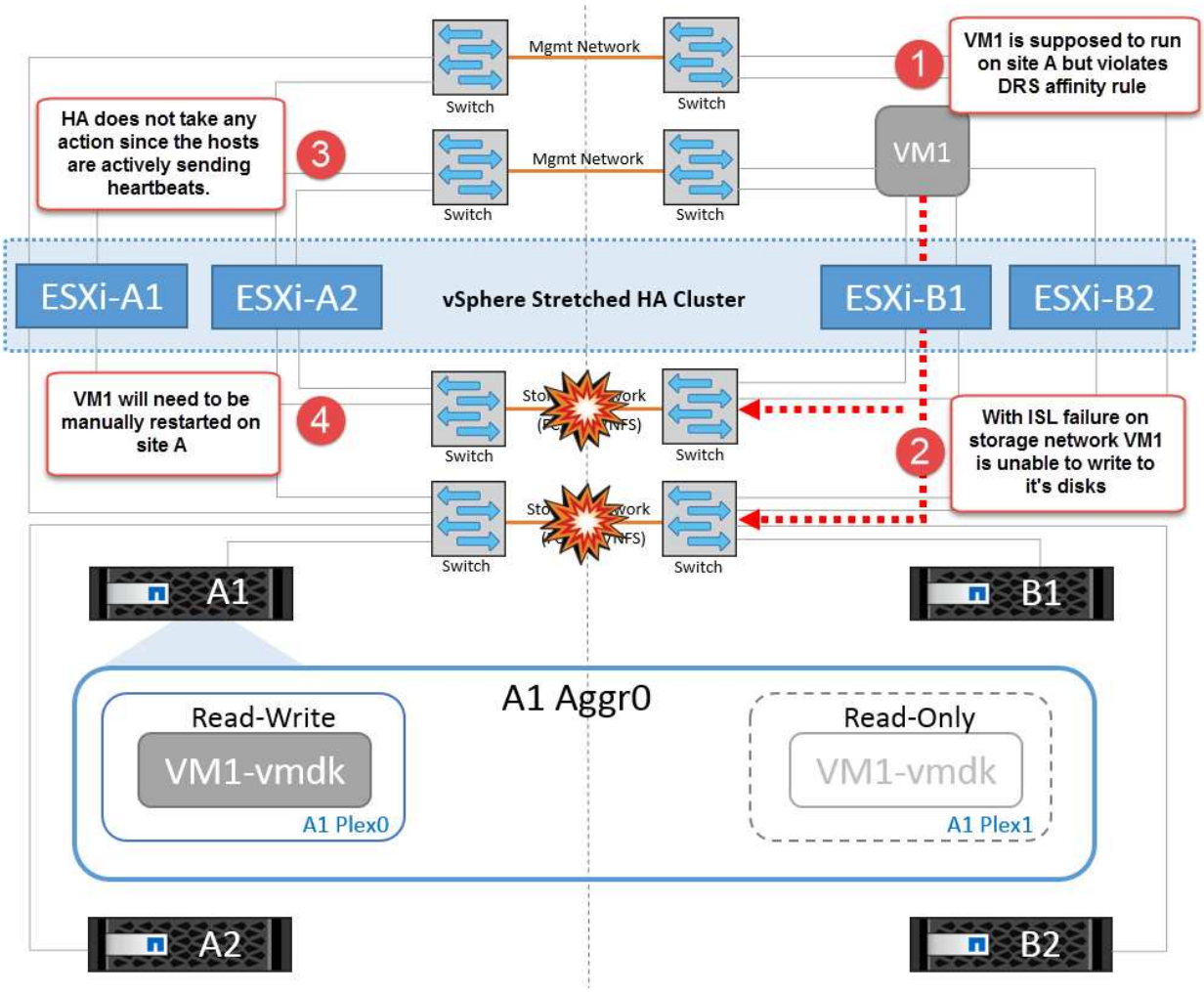


Dans ce scénario, si les liaisons ISL du réseau de stockage back-end tombent en panne, les hôtes du site A perdront l'accès aux volumes de stockage ou aux LUN du cluster B sur le site B et vice versa. Les règles VMware DRS sont définies de manière à ce que l'affinité entre l'hôte et le site de stockage facilite l'exécution des machines virtuelles sans impact sur le site.

Pendant cette période, les machines virtuelles restent en cours d'exécution sur leurs sites respectifs et le comportement de MetroCluster n'a pas changé dans ce scénario. Tous les datastores sont toujours intacts sur leurs sites respectifs.

Si, pour une raison quelconque, la règle d'affinité a été enfreinte (par exemple, VM1, qui était censé s'exécuter à partir du site A où ses disques résident sur les nœuds du cluster A local, s'exécute sur un hôte du site B), le disque de la machine virtuelle est accessible à distance via des liens ISL. En raison d'une défaillance de la

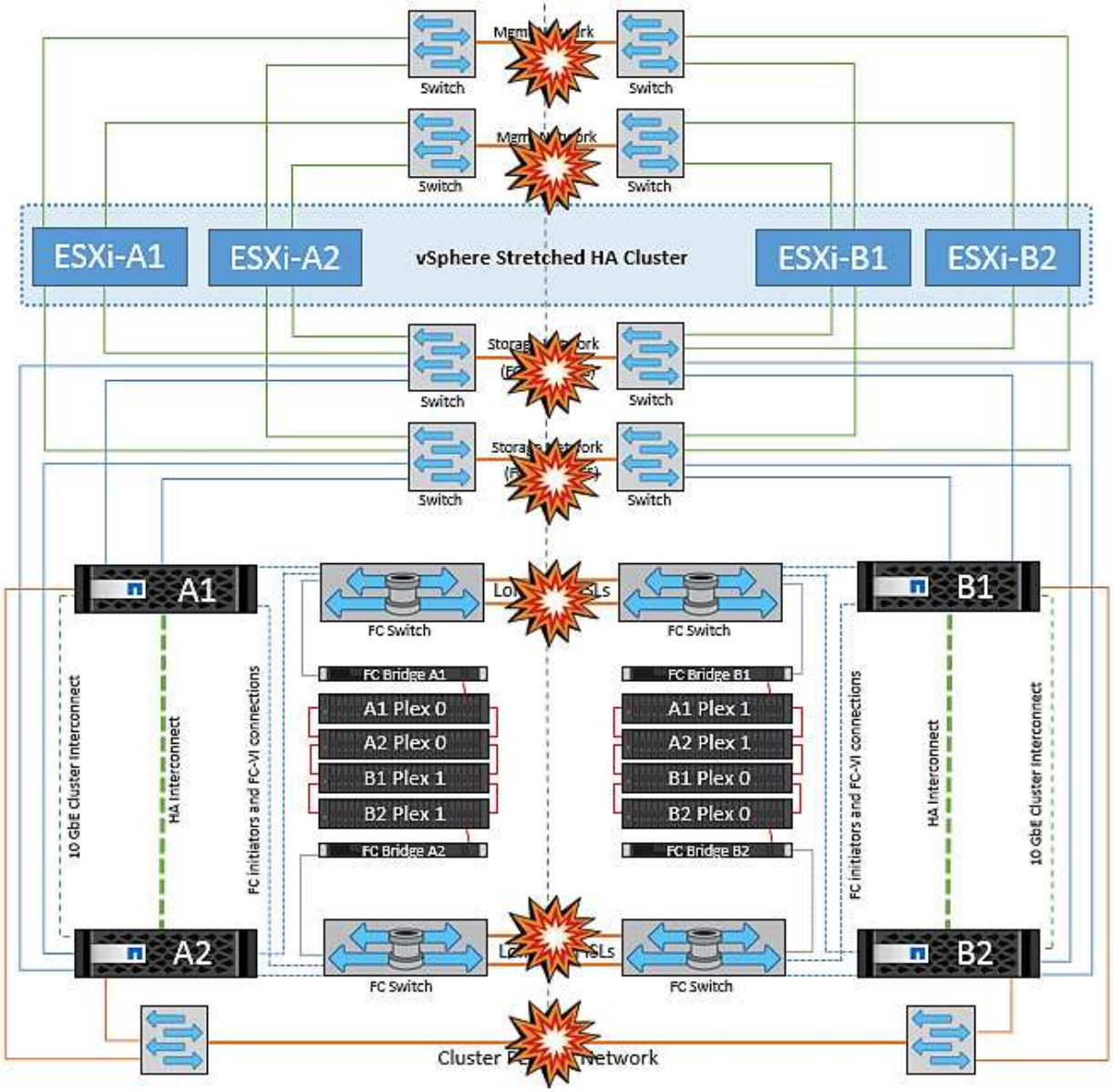
liaison ISL, VM1 exécuté sur le site B ne pouvait pas écrire sur ses disques, car les chemins vers le volume de stockage sont en panne et cette machine virtuelle est en panne. Dans ce cas, VMware HA ne prend aucune action, car les hôtes envoient activement des battements du cœur. Ces machines virtuelles doivent être manuellement désactivées et activées sur leurs sites respectifs. La figure suivante illustre une machine virtuelle violant une règle d'affinité DRS.



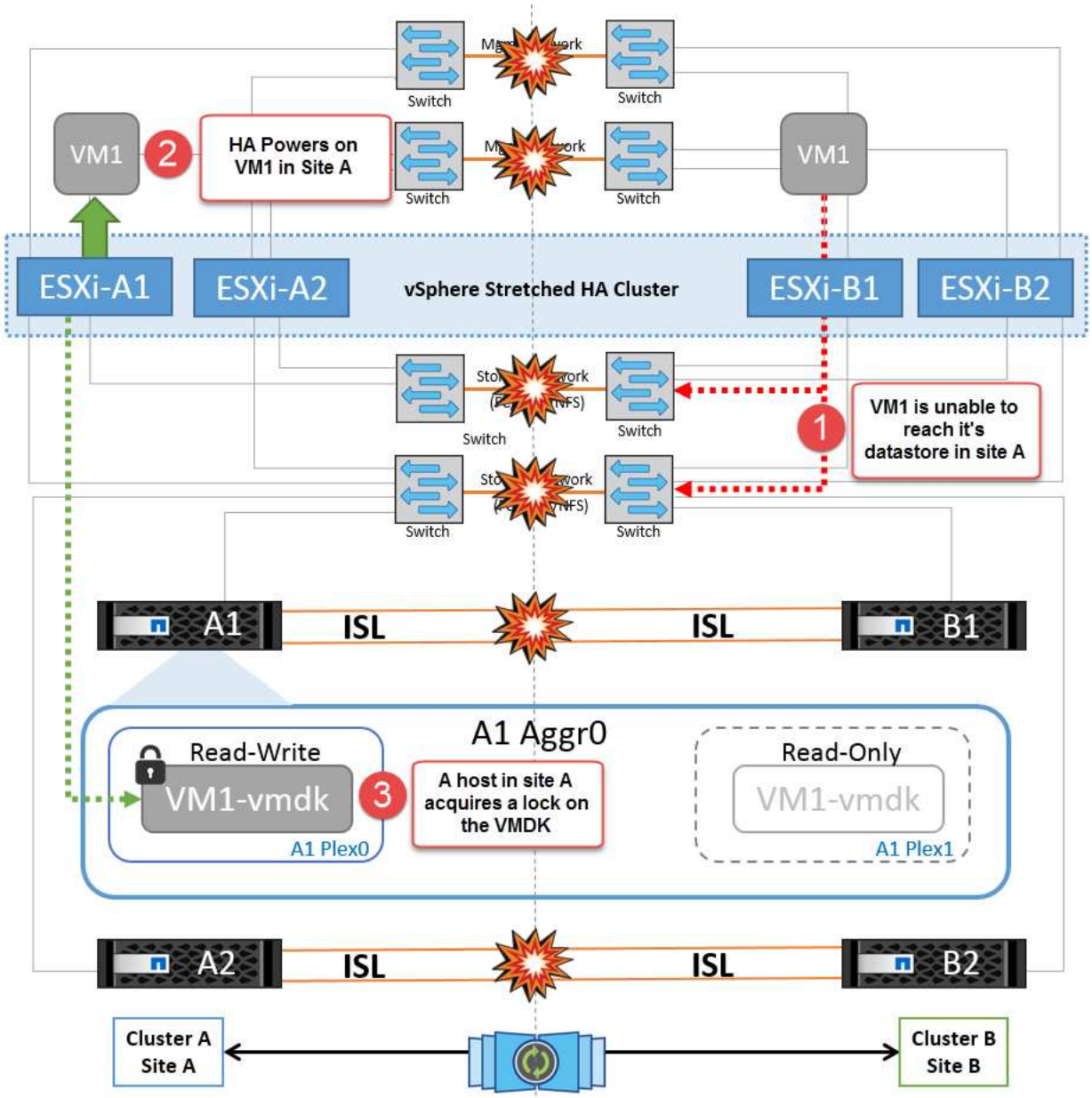
Défaillance de tous les commutateurs ou partition complète du centre de données

Dans ce scénario, toutes les liaisons ISL entre les sites sont en panne et les deux sites sont isolés les uns des autres. Comme nous l'avons vu dans les scénarios précédents, tels que la défaillance des liens ISL au niveau du réseau de gestion et du réseau de stockage, les machines virtuelles ne sont pas affectées par la défaillance complète des liens ISL.

Une fois les hôtes ESXi partitionnés entre les sites, l'agent vSphere HA vérifie la présence de battements de cœur du datastore et, sur chaque site, les hôtes ESXi locaux pourront mettre à jour les battements de cœur du datastore vers leur volume/LUN de lecture/écriture respectif. Les hôtes du site A partent du principe que les autres hôtes ESXi du site B ont échoué car il n'y a pas de pulsations de réseau/datastore. VSphere HA sur le site A tentera de redémarrer les machines virtuelles du site B, ce qui finira par échouer car les datastores du site B ne seront pas accessibles en raison d'une panne de lien ISL du stockage. Une situation similaire est répétée sur le site B.



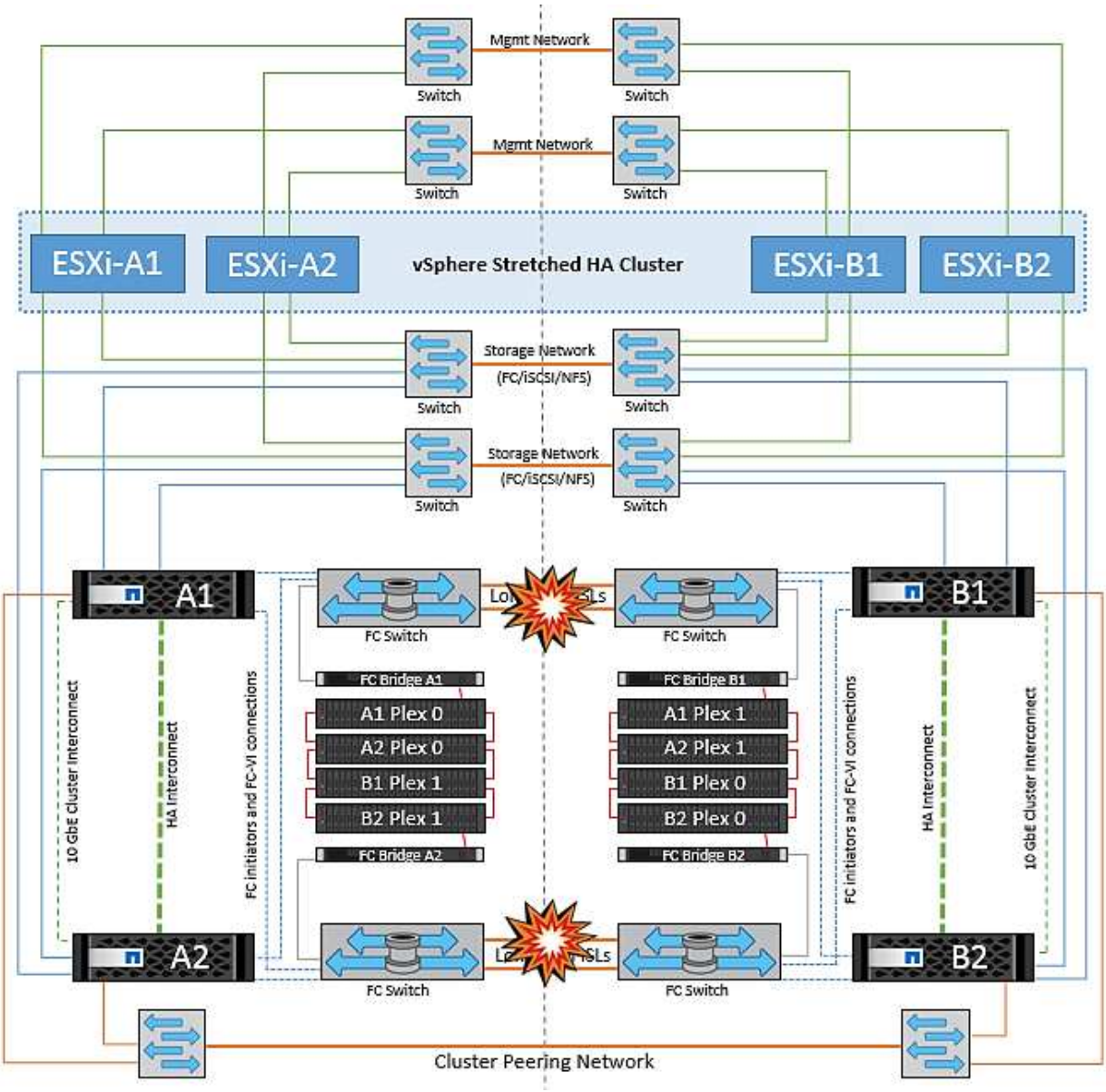
NetApp recommande de déterminer si une machine virtuelle a enfreint les règles DRS. Toutes les machines virtuelles exécutées à partir d'un site distant sont en panne, car elles ne pourront pas accéder au datastore. VSphere HA redémarrera cette machine virtuelle sur le site local. Une fois les liens ISL de nouveau en ligne, la machine virtuelle qui s'exécutait sur le site distant est arrêtée, car il ne peut pas y avoir deux instances de machines virtuelles fonctionnant avec les mêmes adresses MAC.



Défaillance de la liaison inter-commutateur sur les deux fabriques dans NetApp MetroCluster

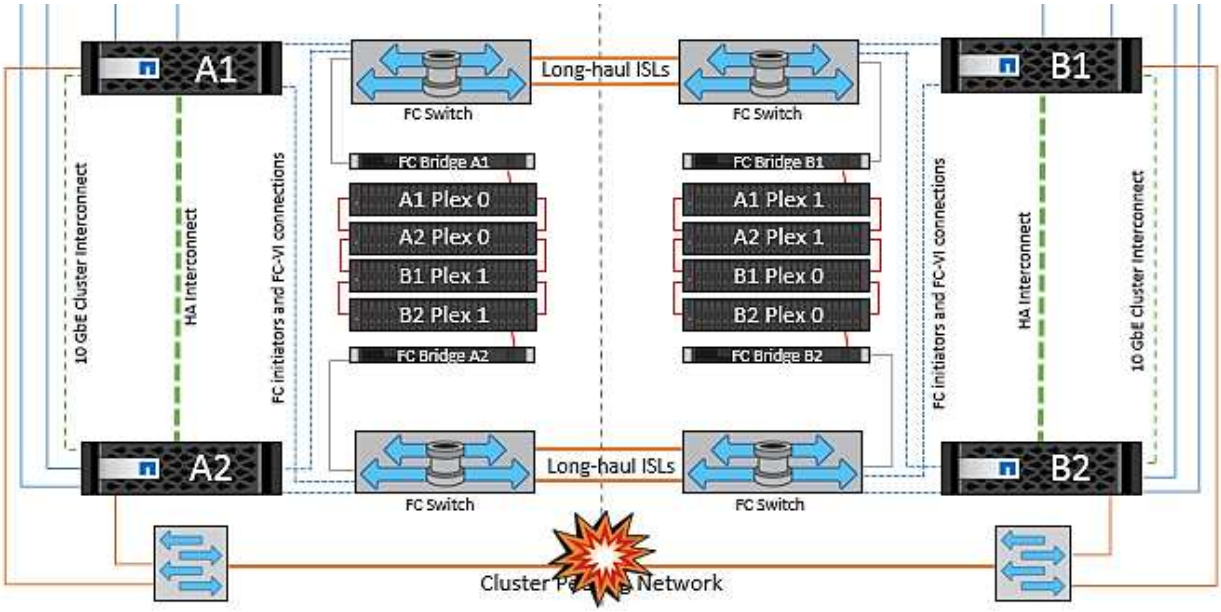
Dans le cas d'une défaillance d'un ou de plusieurs liens ISL, le trafic continue à travers les liens restants. Si toutes les liaisons ISL des deux structures échouent, de sorte qu'il n'y ait pas de liaison entre les sites pour le stockage et la réplication NVRAM, chaque contrôleur continue de transmettre ses données locales. Sur au moins un lien ISL est restauré, la resynchronisation de tous les plexes se produit automatiquement.

Toute écriture effectuée après l'arrêt de toutes les ISL ne sera pas mise en miroir sur l'autre site. Un basculement sur incident, dans cet état, entraînerait la perte des données non synchronisées. Dans ce cas, une intervention manuelle est requise pour la restauration après le basculement. S'il est probable qu'aucune ISL ne soit disponible pendant une période prolongée, l'administrateur peut choisir de fermer tous les services de données afin d'éviter tout risque de perte de données en cas de basculement en cas d'incident. L'exécution de cette action doit être comparée à la probabilité d'un incident nécessitant un basculement avant qu'au moins un lien ISL ne soit disponible. Sinon, si les liens ISL échouent dans un scénario en cascade, un administrateur peut déclencher un basculement planifié vers l'un des sites avant que tous les liens n'aient échoué.



Défaillance du lien de peering de cluster

Dans le cas d'une défaillance de liaison de cluster peering, les liens ISL de la structure sont toujours actifs, les services de données (lectures et écritures) continuent sur les deux sites vers les deux plexes. Toute modification de la configuration du cluster (par exemple, ajout d'un SVM, provisionnement d'un volume ou d'une LUN dans un SVM existant) ne peut pas être propagée à l'autre site. Ils sont conservés dans les volumes de métadonnées CRS locaux et automatiquement propagés à l'autre cluster lors de la restauration de la liaison de cluster peering. Si un basculement forcé est nécessaire avant la restauration de la liaison de cluster peering, les modifications de la configuration du cluster en attente seront automatiquement lues à partir de la copie répliquée à distance des volumes de métadonnées sur le site survivant dans le cadre du processus de basculement.



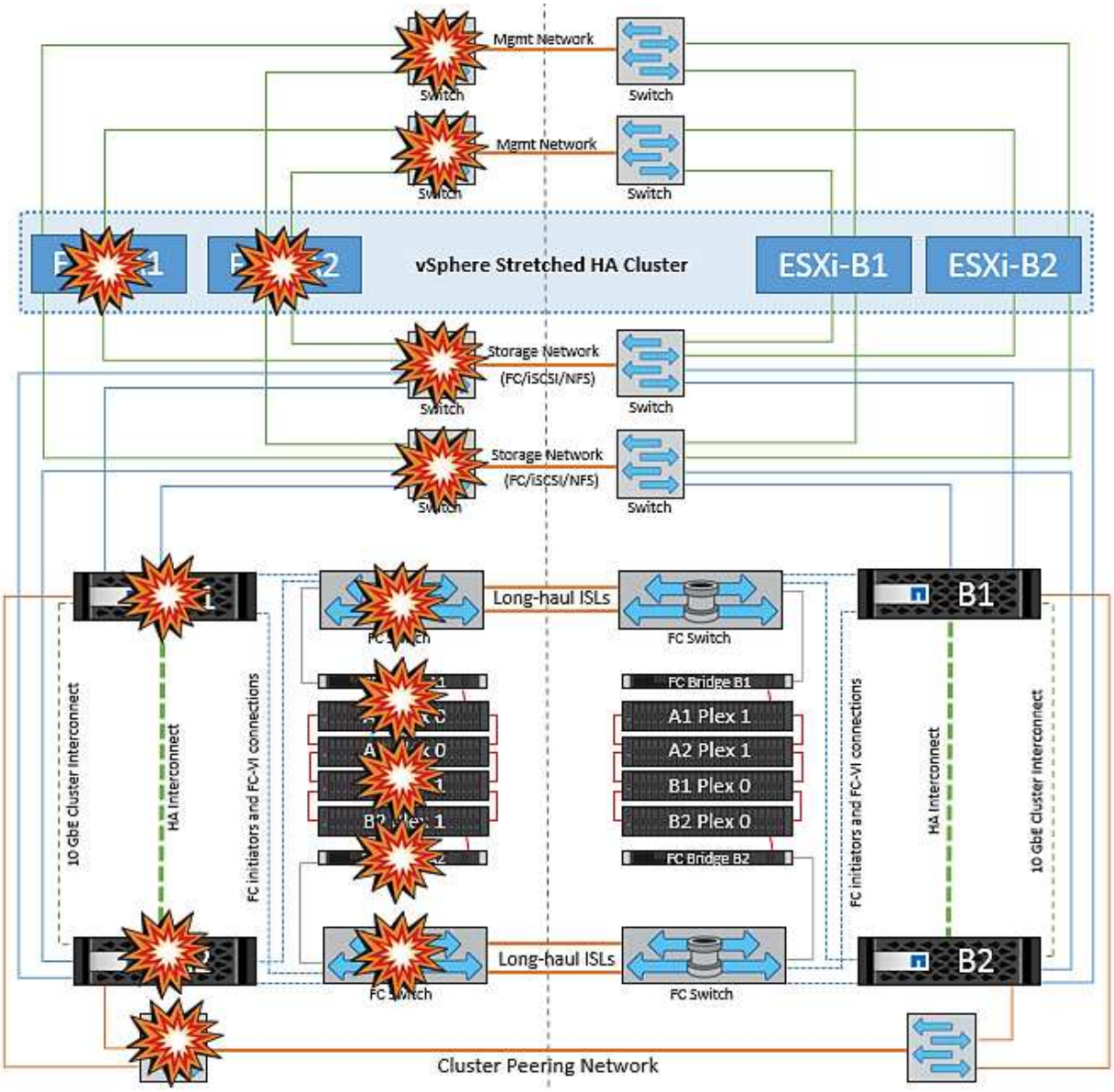
Défaillance complète du site

Dans un scénario de défaillance de site complet A, les hôtes ESXi du site B n'obtiennent pas la pulsation réseau des hôtes ESXi du site A car ils sont en panne. Le maître haute disponibilité sur le site B vérifie que les pulsations du datastore ne sont pas présentes, déclare que les hôtes du site A sont en panne et tente de redémarrer le site A des machines virtuelles sur le site B. Pendant cette période, l'administrateur du stockage effectue un basculement pour reprendre les services des nœuds défaillants sur le site survivant, ce qui restaure tous les services de stockage du site A sur le site B. Une fois que les volumes ou les LUN du site A sont disponibles sur le site B, l'agent principal de haute disponibilité tente de redémarrer le site A des machines virtuelles sur le site B.

Si la tentative de redémarrage d'une machine virtuelle par l'agent principal vSphere HA (qui implique son enregistrement et sa mise sous tension) échoue, le redémarrage est relancé après un délai. Le délai entre les redémarrages peut être configuré jusqu'à un maximum de 30 minutes. VSphere HA tente ces redémarrages au maximum pour un nombre maximal de tentatives (six tentatives par défaut).

> Le maître haute disponibilité ne lance pas les tentatives de redémarrage tant que le gestionnaire des placements n'a pas trouvé le stockage approprié. Dans le cas d'une défaillance complète du site, cela reviendrait à une fois le basculement effectué.

Si le site A été basculé, la panne suivante de l'un des nœuds du site B survivant peut être gérée de manière transparente par le basculement vers le nœud survivant. Dans ce cas, le travail de quatre nœuds est désormais effectué par un seul nœud. Dans ce cas, la restauration consiste à effectuer un rétablissement vers le nœud local. Ensuite, lorsque le site A est restauré, une opération de rétablissement est effectuée pour restaurer le fonctionnement en état stable de la configuration.



Sécurité des produits

Les outils ONTAP pour VMware vSphere

L'ingénierie logicielle avec les outils ONTAP pour VMware vSphere utilise les activités de développement sécurisé suivantes :

- **Modélisation des menaces.** le but de la modélisation des menaces est de découvrir des défauts de sécurité dans une fonction, un composant ou un produit au début du cycle de vie du développement logiciel. Un modèle de menace est une représentation structurée de toutes les informations qui affectent la sécurité d'une application. En substance, c'est une vision de l'application et de son environnement par le biais du principe de sécurité.
- **Dynamic application Security Testing (DAST).** cette technologie est conçue pour détecter les conditions vulnérables sur les applications dans leur état d'exécution. DAST teste les interfaces HTTP et HTML exposées des applications Web-enable.
- **Devise de code tierce.** dans le cadre du développement de logiciels avec des logiciels open-source (OSS), vous devez corriger les vulnérabilités de sécurité qui pourraient être associées à tout OSS intégré à

votre produit. Il s'agit d'un effort continu car une nouvelle version OSS peut avoir une nouvelle vulnérabilité découverte signalée à tout moment.

- **Analyse des vulnérabilités** l'analyse des vulnérabilités a pour but de détecter les vulnérabilités de sécurité courantes et connues dans les produits NetApp avant leur publication auprès des clients.
- * Tests de pénétration.* le test de pénétration est le processus d'évaluation d'un système, d'une application Web ou d'un réseau pour trouver des vulnérabilités de sécurité qui pourraient être exploitées par un attaquant. Les tests d'intrusion chez NetApp sont réalisés par un groupe d'entreprises tierces de confiance et approuvées. Leur domaine de test comprend le lancement d'attaques contre une application ou un logiciel similaire à des intrus hostiles ou des pirates informatiques à l'aide de méthodes ou d'outils d'exploitation sophistiqués.

Fonctionnalités de sécurité du produit

Les outils ONTAP pour VMware vSphere comprennent les fonctions de sécurité suivantes dans chaque version.

- **Bannière de connexion.** SSH est désactivé par défaut et n'autorise que les connexions à une seule fois si elles sont activées à partir de la console VM. La bannière de connexion suivante s'affiche une fois que l'utilisateur a saisi un nom d'utilisateur dans l'invite de connexion :

  AVERTISSEMENT: l'accès non autorisé à ce système est interdit et sera poursuivi par la loi. En accédant à ce système, vous convenez que vos actions peuvent être surveillées si vous soupçonnez une utilisation non autorisée.

  Une fois la connexion établie par l'utilisateur via le canal SSH, le texte suivant s'affiche :

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- **Contrôle d'accès basé sur les rôles (RBAC).** deux types de contrôles RBAC sont associés aux outils ONTAP :
  - Privilèges de serveur vCenter natif
  - Privilèges spécifiques au plug-in vCenter. Pour plus de détails, voir "ce lien".
- **Canaux de communication cryptés.** toutes les communications externes se produisent sur HTTPS en utilisant la version 1.2 de TLS.
- **Exposition minimale au port.** seuls les ports nécessaires sont ouverts sur le pare-feu.

  Le tableau suivant décrit les détails du port ouvert.

N° de port TCP v4/v6	Direction	Fonction
8143	entrant	Connexions HTTPS pour l'API REST
8043	entrant	Connexions HTTPS

N° de port TCP v4/v6	Direction	Fonction
9060	entrant	Connexions HTTPS Utilisé pour les connexions SOAP sur https Ce port doit être ouvert pour permettre à un client de se connecter au serveur d'API des outils ONTAP.
22	entrant	SSH (désactivé par défaut)
9080	entrant	Connexions HTTPS - VP et SRA - connexions internes à partir du bouclage uniquement
9083	entrant	Connexions HTTPS - VP et SRA Utilisé pour les connexions SOAP sur https
1162	entrant	Paquets de déroutement SNMP VP
1527	diffusion interne uniquement	Port de base de données Derby, uniquement entre cet ordinateur et lui-même, connexions externes non acceptées — connexions internes uniquement
443	bidirectionnel	Utilisé pour les connexions aux clusters ONTAP

- **Prise en charge des certificats signés de l'autorité de certification (CA).** les outils ONTAP pour VMware vSphere prennent en charge les certificats signés de l'autorité de certification. Voir ceci "article de la base de connaissances" pour en savoir plus.
- **Audit Logging.** les offres de support peuvent être téléchargées et sont extrêmement détaillées. Les outils ONTAP consigne toutes les activités de connexion et de déconnexion de l'utilisateur dans un fichier journal distinct. Les appels d'API VASA sont connectés à un journal d'audit VASA dédié (local cxf.log).
- **Stratégies de mot de passe.** les stratégies de mot de passe suivantes sont respectées :
  - Les mots de passe ne sont pas enregistrés dans des fichiers journaux.
  - Les mots de passe ne sont pas communiqués en texte brut.
  - Les mots de passe sont configurés lors du processus d'installation lui-même.
  - L'historique des mots de passe est un paramètre configurable.
  - L'âge minimum du mot de passe est défini sur 24 heures.
  - La saisie automatique des champs de mot de passe est désactivée.
  - Les outils ONTAP crypte toutes les informations d'identification stockées à l'aide de la fonction de hachage SHA256.

Plug-in SnapCenter VMware vSphere

Le plug-in NetApp SnapCenter pour l'ingénierie logicielle VMware vSphere exploite les activités de développement sécurisées suivantes :

- **Modélisation des menaces.** le but de la modélisation des menaces est de découvrir des défauts de sécurité dans une fonction, un composant ou un produit au début du cycle de vie du développement logiciel. Un modèle de menace est une représentation structurée de toutes les informations qui affectent la sécurité d'une application. En substance, c'est une vision de l'application et de son environnement par le biais du principe de sécurité.
- **Test dynamique de sécurité des applications (DAST).** technologies conçues pour détecter les conditions vulnérables des applications dans leur état d'exécution. DAST teste les interfaces HTTP et HTML exposées des applications Web-enable.
- **Devise de code tierce.** dans le cadre du développement de logiciels et de l'utilisation de logiciels open-source (OSS), il est important de traiter les vulnérabilités de sécurité qui pourraient être associées à OSS qui a été intégré à votre produit. Il s'agit d'un effort continu car la version du composant OSS peut avoir une vulnérabilité nouvellement découverte signalée à tout moment.
- **Analyse des vulnérabilités** l'analyse des vulnérabilités a pour but de détecter les vulnérabilités de sécurité courantes et connues dans les produits NetApp avant leur publication auprès des clients.
- * Tests de pénétration.* le test de pénétration est le processus d'évaluation d'un système, d'une application Web ou d'un réseau pour trouver des vulnérabilités de sécurité qui pourraient être exploitées par un attaquant. Les tests d'intrusion chez NetApp sont réalisés par un groupe d'entreprises tierces de confiance et approuvées. Leur domaine de test comprend le lancement d'attaques contre une application ou un logiciel comme des intrus hostiles ou des pirates informatiques à l'aide de méthodes ou d'outils d'exploitation sophistiqués.
- **Activité de réponse aux incidents de sécurité des produits.** les vulnérabilités de sécurité sont découvertes à la fois en interne et en externe dans l'entreprise et peuvent constituer un risque sérieux pour la réputation de NetApp si elles ne sont pas traitées dans les délais impartis. Pour faciliter ce processus, l'équipe d'intervention en cas d'incident de sécurité des produits (PSIRT) signale et effectue le suivi des vulnérabilités.

Fonctionnalités de sécurité du produit

Le plug-in NetApp SnapCenter pour VMware vSphere inclut les fonctionnalités de sécurité suivantes dans chaque version :

- **Accès limité au shell.** SSH est désactivé par défaut, et les connexions à une seule fois ne sont autorisées que si elles sont activées à partir de la console VM.
- **Avertissement d'accès dans la bannière de connexion.** la bannière de connexion suivante s'affiche après que l'utilisateur ait entré un nom d'utilisateur dans l'invite de connexion :

  AVERTISSEMENT: l'accès non autorisé à ce système est interdit et sera poursuivi par la loi. En accédant à ce système, vous convenez que vos actions peuvent être surveillées si vous soupçonnez une utilisation non autorisée.

  Une fois que l'utilisateur a terminé sa connexion via le canal SSH, les valeurs de sortie suivantes s'affichent :

```
Linux vsc1 4.19.0-12-amd64 #1 SMP Debian 4.19.152-1 (2020-10-18) x86_64
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

- **Contrôle d'accès basé sur les rôles (RBAC).** deux types de contrôles RBAC sont associés aux outils ONTAP :
  - Privilèges de serveur vCenter natif.
  - Privilèges spécifiques au plug-in VMware vCenter. Pour plus d'informations, voir "Contrôle d'accès basé sur des rôles (RBAC)".
- **Canaux de communication cryptés.** toutes les communications externes sont effectuées via HTTPS en utilisant TLS.
- **Exposition minimale au port.** seuls les ports nécessaires sont ouverts sur le pare-feu.

Le tableau suivant fournit les détails du port ouvert.

Numéro de port TCP v4/v6	Fonction
8144	Connexions HTTPS pour l'API REST
8080	Connexions HTTPS pour interface graphique OVA
22	SSH (désactivé par défaut)
3306	MySQL (connexions internes uniquement, connexions externes désactivées par défaut)
443	Nginx (services de protection des données)

- **Prise en charge des certificats signés par l'autorité de certification (CA).** le plug-in SnapCenter pour VMware vSphere prend en charge la fonctionnalité des certificats signés par l'autorité de certification. Voir "Comment créer et/ou importer un certificat SSL dans le plug-in SnapCenter pour VMware vSphere (SCV)".
- **Stratégies de mot de passe.** les stratégies de mot de passe suivantes sont en vigueur :
  - Les mots de passe ne sont pas enregistrés dans des fichiers journaux.
  - Les mots de passe ne sont pas communiqués en texte brut.
  - Les mots de passe sont configurés lors du processus d'installation lui-même.
  - Toutes les informations d'identification sont stockées à l'aide d'un hachage SHA256.
- **Image du système d'exploitation de base.** le produit est fourni avec le système d'exploitation de base Debian pour OVA avec accès restreint et accès au shell désactivé. Cela réduit l'empreinte d'attaque. Chaque système d'exploitation de base SnapCenter est mis à jour avec les derniers correctifs de sécurité disponibles pour une protection maximale.

NetApp développe des fonctionnalités logicielles et des correctifs de sécurité en ce qui concerne le plug-in SnapCenter pour l'appliance VMware vSphere, puis les publie auprès de ses clients sous la forme d'un pack logiciel. Étant donné que ces dispositifs intègrent des dépendances spécifiques au système d'exploitation Linux et à notre logiciel propriétaire, NetApp vous recommande de ne pas modifier le système sous-exploitation, car il présente un potentiel important d'affecter l'appliance NetApp. Cela pourrait affecter la capacité de NetApp à prendre en charge l'appliance. NetApp recommande de tester et de déployer la dernière version de code pour les appliances, car elles sont publiées pour corriger les problèmes de sécurité.

Guide de renforcement de la sécurité des outils ONTAP pour VMware vSphere

Guide de renforcement de la sécurité des outils ONTAP pour VMware vSphere 9.13

Le guide de renforcement de la sécurité des outils ONTAP pour VMware vSphere fournit un ensemble complet d'instructions pour configurer les paramètres les plus sécurisés.

Ces guides s'appliquent à la fois aux applications et au système d'exploitation invité de l'appliance elle-même.

Vérification de l'intégrité des outils ONTAP pour les packages d'installation de VMware vSphere 9.13

Deux méthodes sont disponibles pour vérifier l'intégrité des packages d'installation des outils ONTAP.

1. Vérification des checksums
2. Vérification de la signature

Les sommes de contrôle sont fournies sur les pages de téléchargement des paquets d'installation d'OTV. Les utilisateurs doivent vérifier les sommes de contrôle des paquets téléchargés par rapport à la somme de contrôle fournie sur la page de téléchargement.

Vérification de la signature des outils ONTAP OVA

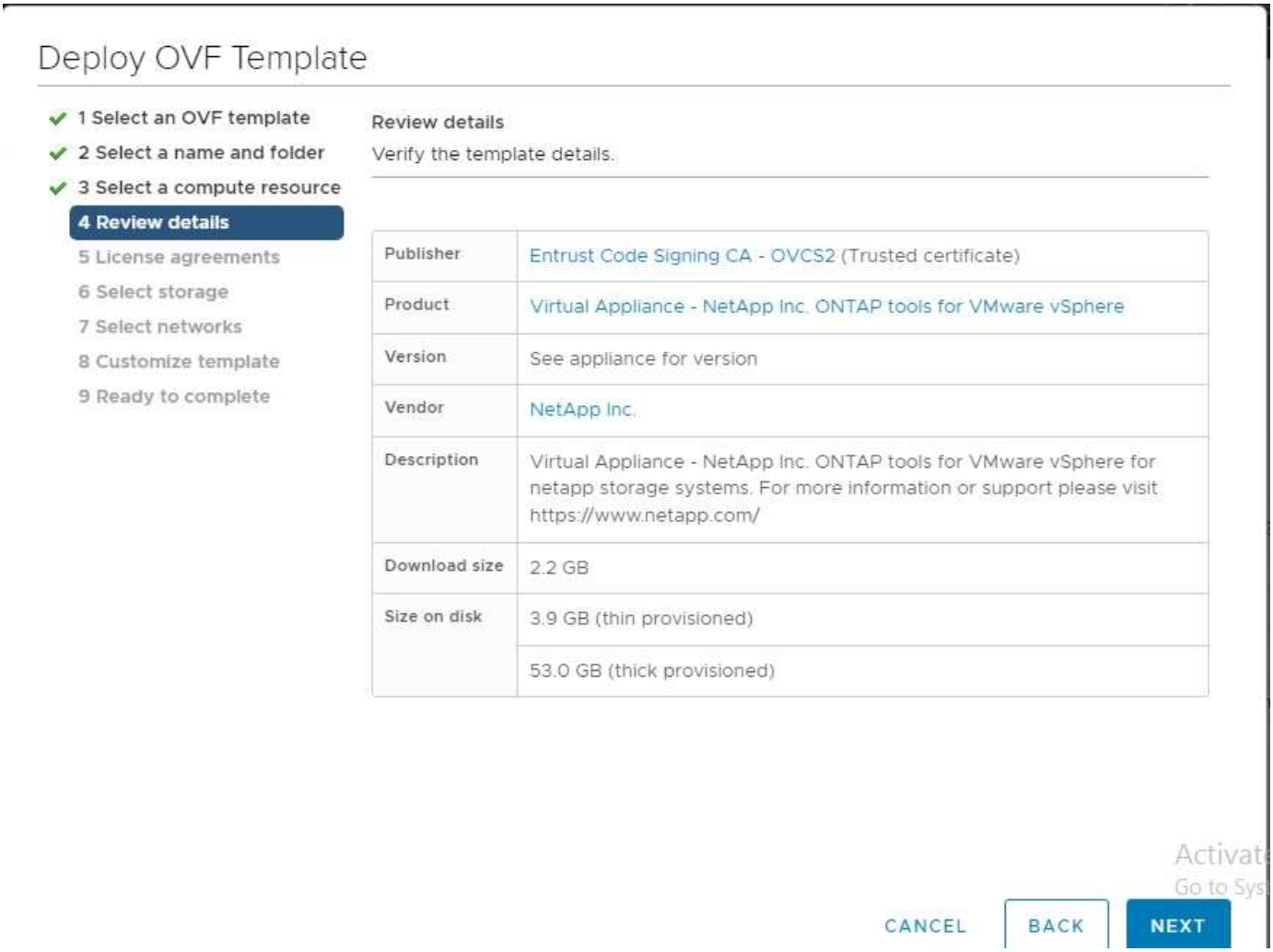
Le paquet d'installation de vApp est livré sous la forme d'une boule de commande. Ce tarball contient des certificats intermédiaires et racine pour l'appliance virtuelle, ainsi qu'un fichier README et un package OVA. Le fichier README guide les utilisateurs sur la façon de vérifier l'intégrité du progiciel VApp OVA.

Les clients doivent également télécharger les certificats racine et intermédiaire fournis sur vCenter version 7.0U3E et ultérieure. Pour les versions vCenter comprises entre 7.0.1 et 7.0.U3E, la fonctionnalité de vérification du certificat n'est pas prise en charge par VMware. Les clients n'ont pas besoin de télécharger de certificat pour vCenter versions 6.x.

Téléchargement du certificat racine sécurisé vers vCenter

1. Connectez-vous à vCenter Server à l'aide du client VMware vSphere.
2. Spécifiez le nom d'utilisateur et le mot de passe de laman@vspher.locl ou d'un autre membre du groupe administrateurs d'authentification unique vCenter. Si vous avez spécifié un domaine différent lors de l'installation, connectez-vous en tant qu'administrateur@mondomaine.
3. Accédez à l'interface utilisateur de gestion des certificats : a. dans le menu Accueil, sélectionnez Administration. b. sous certificats, cliquez sur gestion des certificats.
4. Si le système vous y invite, entrez les informations d'identification de votre serveur vCenter.
5. Sous certificats racine approuvés, cliquez sur Ajouter.
6. Cliquez sur Parcourir et sélectionnez l'emplacement du fichier .pem du certificat (OTV_OVA_INTER_ROOT_CERT_CHAIN.pem).
7. Cliquez sur Ajouter. Le certificat est ajouté au magasin.

Reportez-vous à la section "Ajoutez un certificat racine de confiance au magasin de certificats" pour en savoir plus. Lors du déploiement d'une vApp (à l'aide du fichier OVA), la signature numérique du package vApp peut être vérifiée sur la page « Review details » (vérifier les détails). Si le package vApp téléchargé est authentique, la colonne « Éditeur » affiche « certificat de confiance » (comme dans la capture d'écran suivante).



Vérification de la signature des outils ONTAP ISO et SRA tar.gz

NetApp partage son certificat de signature de code avec les clients sur la page de téléchargement du produit, ainsi que les fichiers zip du produit pour OTV-ISO et SRA.tgz.

À partir du certificat de signature de code, les utilisateurs peuvent extraire la clé publique comme suit :

```
 #> openssl x509 -in <code-sign-cert, pem file> -pubkey -noout > <public-
key name>
Ensuite, la clé publique doit être utilisée pour vérifier la signature
pour iso et tgz produit zip comme ci-dessous :
```

```
 #> openssl dgst -sha256 -verify <public-key> -signature <signature-file>
<binary-name>
Exemple :
```

```
#> openssl x509 -in OTV_ISO_CERT.pem -pubkey -noout > OTV_ISO.pub
#> openssl dgst -sha256 -verify OTV_ISO.pub -signature netapp-ontap-tools-
for-vmware-vsphere-9.12-upgrade-iso.sig netapp-ontap-tools-for-vmware-
vsphere-9.12-upgrade.iso
Verified OK  => response
```

Ports et protocoles pour les outils ONTAP 9.13

La liste ci-dessous répertorie les ports et les protocoles requis permettant la communication entre les outils ONTAP pour le serveur VMware vSphere et d'autres entités telles que les systèmes de stockage géré, les serveurs et d'autres composants.

Ports entrants et sortants requis pour OTV

Notez le tableau ci-dessous qui répertorie les ports entrants et sortants requis pour le bon fonctionnement des outils ONTAP. Il est important de s'assurer que seuls les ports mentionnés dans le tableau sont ouverts pour les connexions à partir de machines distantes, tandis que tous les autres ports doivent être bloqués pour les connexions à partir de machines distantes. Cela permet d'assurer la sécurité de votre système.

Le tableau suivant décrit les détails du port ouvert.

Port TCP v4/v6 #	Direction	Fonction
8143	entrant	Connexions HTTPS pour l'API REST
8043	entrant	Connexions HTTPS
9060	entrant	Connexions HTTPS Utilisé pour les connexions SOAP sur HTTPS Ce port doit être ouvert pour permettre à un client de se connecter au serveur d'API des outils ONTAP.
22	entrant	SSH (désactivé par défaut)
9080	entrant	Connexions HTTPS - VP et SRA - connexions internes à partir du bouclage uniquement
9083	entrant	Connexions HTTPS - VP et SRA Utilisé pour les connexions SOAP sur HTTPS
1162	entrant	Paquets de déroutement SNMP VP
8443	entrant	Plug-in distant
1527	diffusion interne uniquement	Port de base de données Derby, uniquement entre cet ordinateur et lui-même, connexions externes non acceptées — connexions internes uniquement
8150	diffusion interne uniquement	Le service d'intégrité des journaux s'exécute sur le port
443	bidirectionnel	Utilisé pour les connexions aux clusters ONTAP

Contrôle de l'accès à distance à la base de données Derby

Les administrateurs peuvent accéder à la base de données derby à l'aide des commandes suivantes. Il est accessible via la machine virtuelle locale des outils ONTAP ainsi qu'un serveur distant en procédant comme suit :

```
java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
connect 'jdbc:derby://<OTV-
IP>:1527//opt/netapp/vpserver/vvoldb;user=<user>;password=<password>';
```

exemple:

```
root@UnifiedVSC:~# java -classpath "/opt/netapp/vpserver/lib/*" org.apache.derby.tools.ij;
ij version 10.15
ij> connect 'jdbc:derby://localhost:1527//opt/netapp/vpserver/vvoldb;user=app;password=[redacted]';
ij> show tables;
TABLE_SCHEM         |TABLE_NAME                    |REMARKS
------------------------------------------------------------------------
SYS                 |SYSALIASES                    |
SYS                 |SYSCHECKS                     |
SYS                 |SYSCOLPERMS                   |
SYS                 |SYSCOLUMNS                    |
SYS                 |SYSCONGLOMERATES              |
SYS                 |SYSCONSTRAINTS                |
SYS                 |SYSDEPENDS                    |
SYS                 |SYSFILES                      |
SYS                 |SYSFOREIGNKEYS                |
SYS                 |SYSKEYS                       |
SYS                 |SYSPERMS                      |
```

Outils ONTAP pour les points d'accès VMware vSphere 9.13 (utilisateurs)

L'installation des outils ONTAP pour VMware vSphere crée et utilise trois types d'utilisateurs :

1. Utilisateur système : compte utilisateur root
2. Utilisateur de l'application : l'utilisateur administrateur, l'utilisateur maint et les comptes utilisateur db
3. Utilisateur de support : compte utilisateur diag

1. Utilisateur du système

L'utilisateur System(root) est créé par l'installation des outils ONTAP sur le système d'exploitation sous-jacent (Debian).

- Un utilisateur système par défaut "root" est créé sur Debian par l'installation des outils ONTAP. Sa valeur par défaut est désactivée et peut être activée ad hoc via la console « maint ».

2. Utilisateur de l'application

L'utilisateur de l'application est nommé en tant qu'utilisateur local dans les outils ONTAP. Il s'agit d'utilisateurs créés dans l'application Outils ONTAP. Le tableau ci-dessous répertorie les types d'utilisateurs d'applications :

Utilisateur	Description
Utilisateur administrateur	Il est créé lors de l'installation des outils ONTAP et l'utilisateur fournit les informations d'identification lors du déploiement des outils ONTAP. Les utilisateurs ont la possibilité de modifier le mot de passe dans la console « maint ». Le mot de passe expirera dans 90 jours et les utilisateurs devraient changer de mot de passe.
Utilisateur de maintenance	Il est créé lors de l'installation des outils ONTAP et l'utilisateur fournit les informations d'identification lors du déploiement des outils ONTAP. Les utilisateurs ont la possibilité de modifier le mot de passe dans la console « maint ». Il s'agit d'un utilisateur de maintenance créé pour exécuter les opérations de la console de maintenance.
Utilisateur de la base de données	Il est créé lors de l'installation des outils ONTAP et l'utilisateur fournit les informations d'identification lors du déploiement des outils ONTAP. Les utilisateurs ont la possibilité de modifier le mot de passe dans la console « maint ». Le mot de passe expirera dans 90 jours et les utilisateurs devraient changer de mot de passe.

3. Support user(diag user)

Lors de l'installation des outils ONTAP, un utilisateur du support est créé. Cet utilisateur peut accéder aux outils ONTAP en cas de problème ou de panne du serveur et collecter les journaux. Par défaut, cet utilisateur est désactivé, mais il peut être activé sur une base ad hoc via la console « maint ». Il est important de noter que cet utilisateur sera automatiquement désactivé après une certaine période.

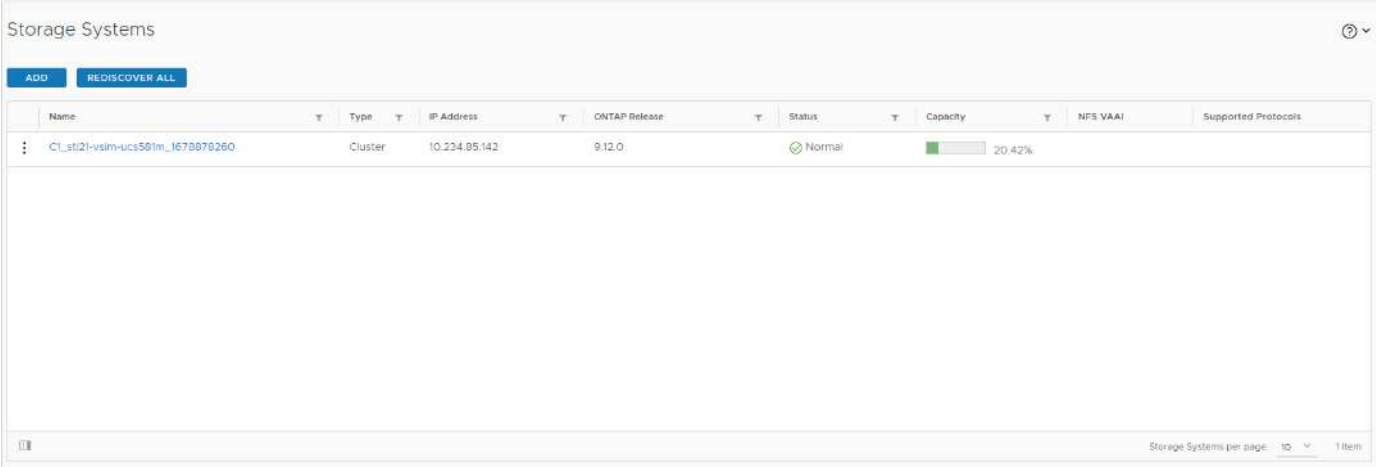
ONTAP Tools 9.13 Mutual TLS (authentification basée sur certificat)

Les versions 9.7 et ultérieures de ONTAP prennent en charge les communications TLS mutuelles. Depuis les outils ONTAP pour VMware et vSphere 9.12, le protocole TLS mutuel est utilisé pour la communication avec les nouveaux clusters ajoutés (selon la version de ONTAP).

ONTAP

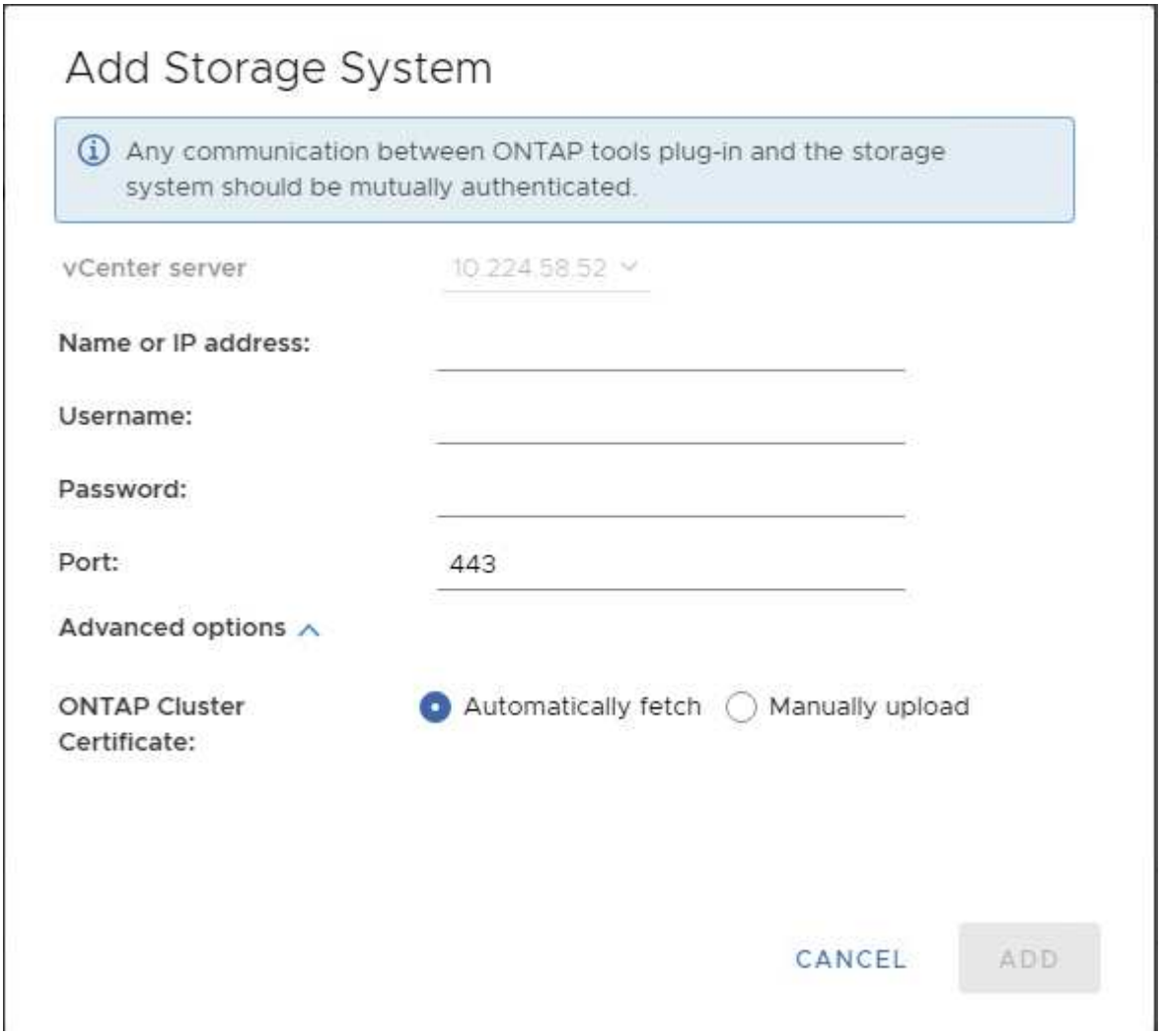
Pour tous les systèmes de stockage précédemment ajoutés : lors d'une mise à niveau, tous les systèmes de stockage ajoutés font l'objet d'une fiabilité automatique et les mécanismes d'authentification basés sur des certificats sont configurés.

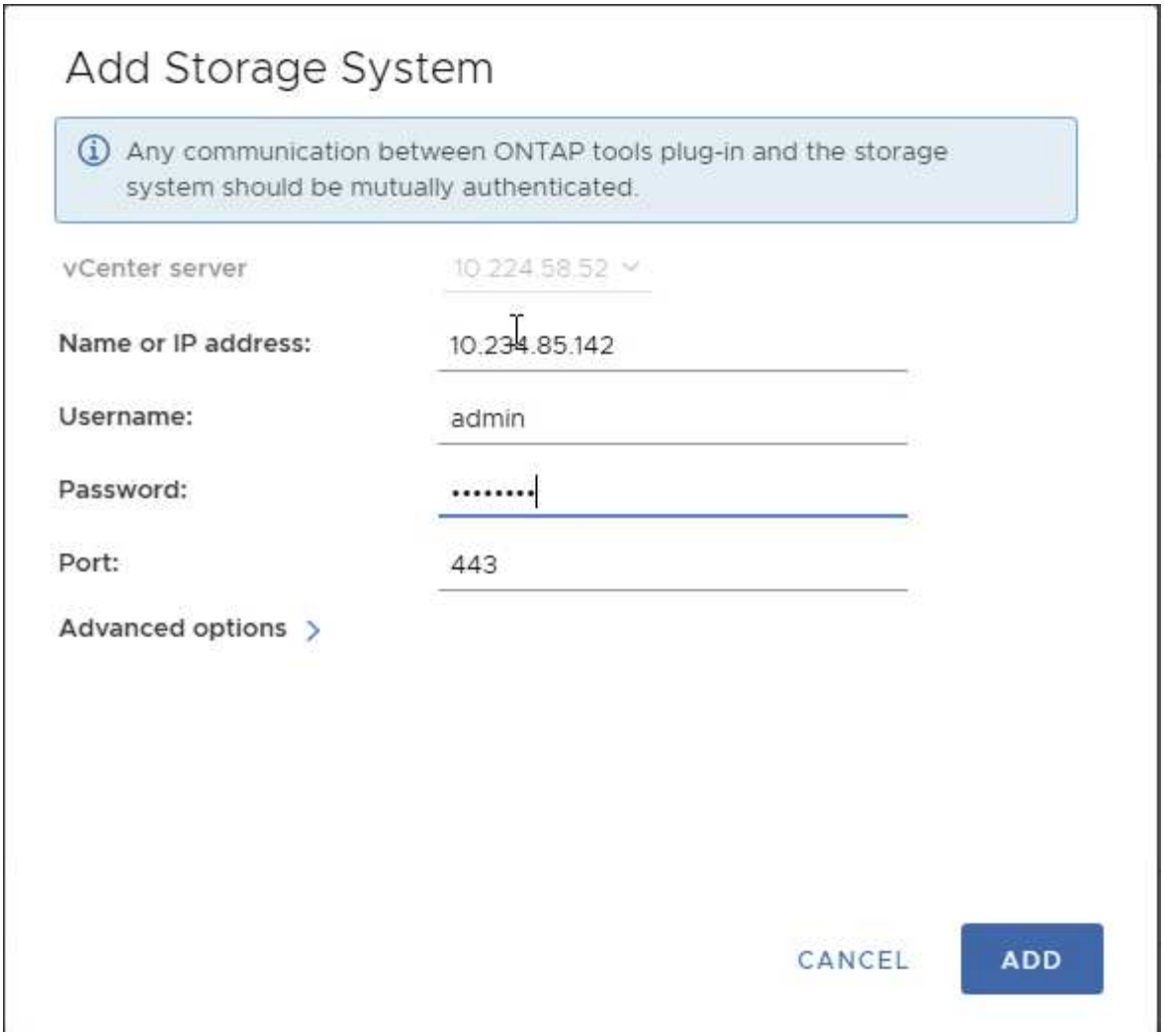
Comme dans la capture d'écran ci-dessous, la page de configuration du cluster affiche l'état d'authentification mutuelle TLS (Certificate Based Authentication), configurée pour chaque cluster.

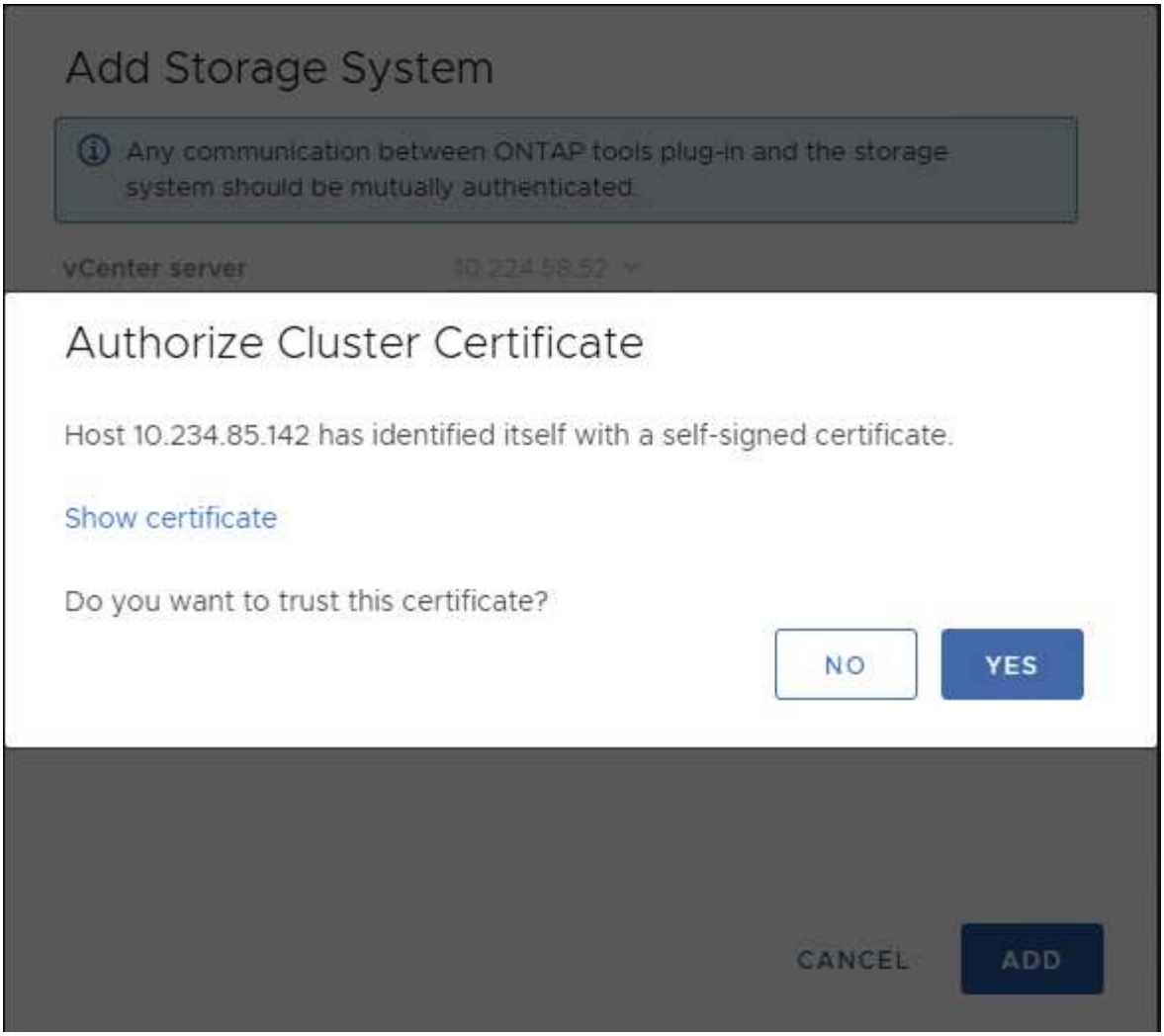


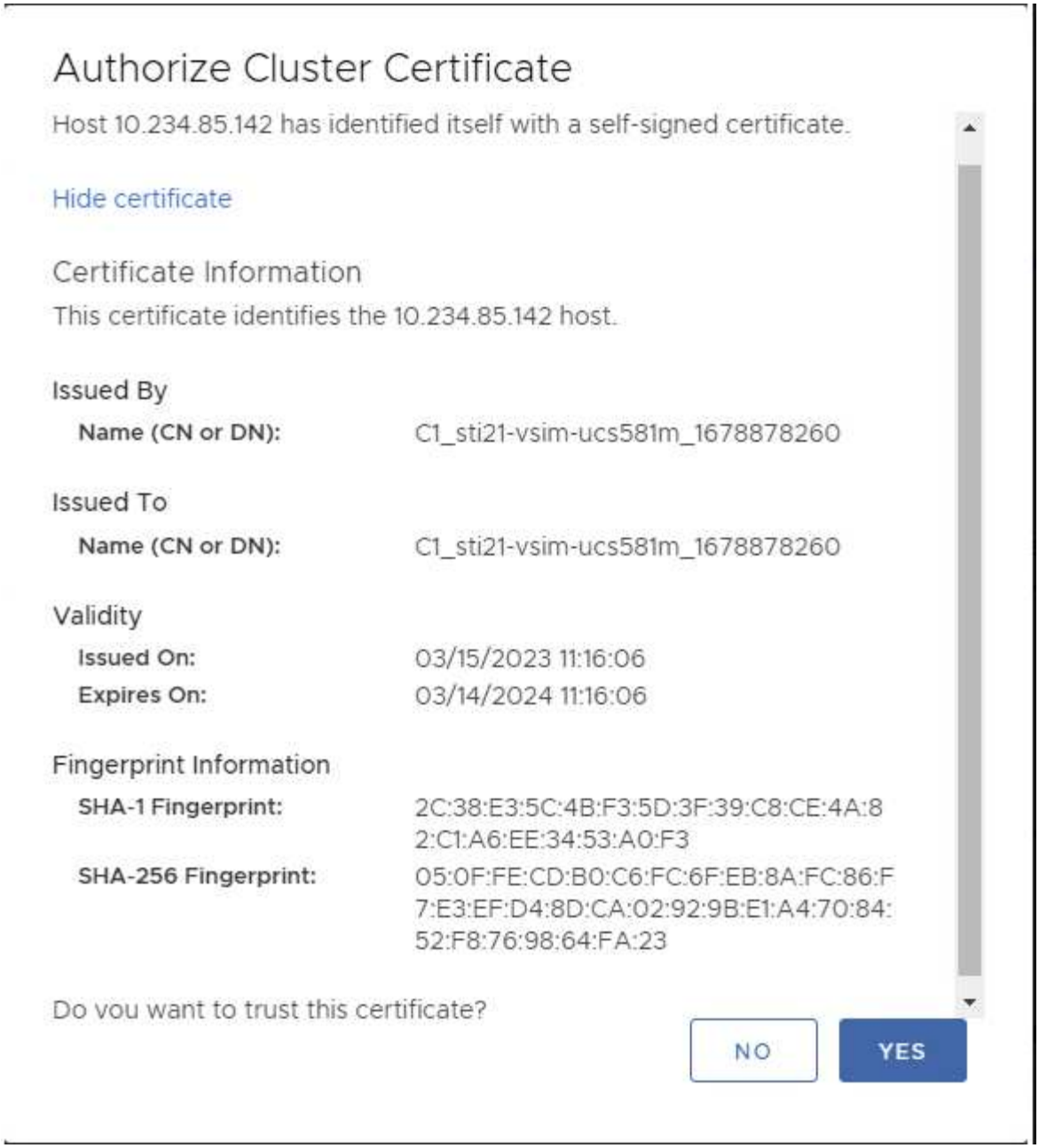
Cluster Add

Lors du workflow d'ajout de cluster, si le cluster ajouté prend en charge MTLS, MTLS sera configuré par défaut. L'utilisateur n'a pas besoin d'effectuer de configuration pour cela. La capture d'écran ci-dessous présente l'écran présenté à l'utilisateur lors de l'ajout d'un cluster.




Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
10.224.58.52
Name or IP address:
10.234.85.142
Username:
admin
Password:
Port:
443
Advanced options
CANCEL
ADD

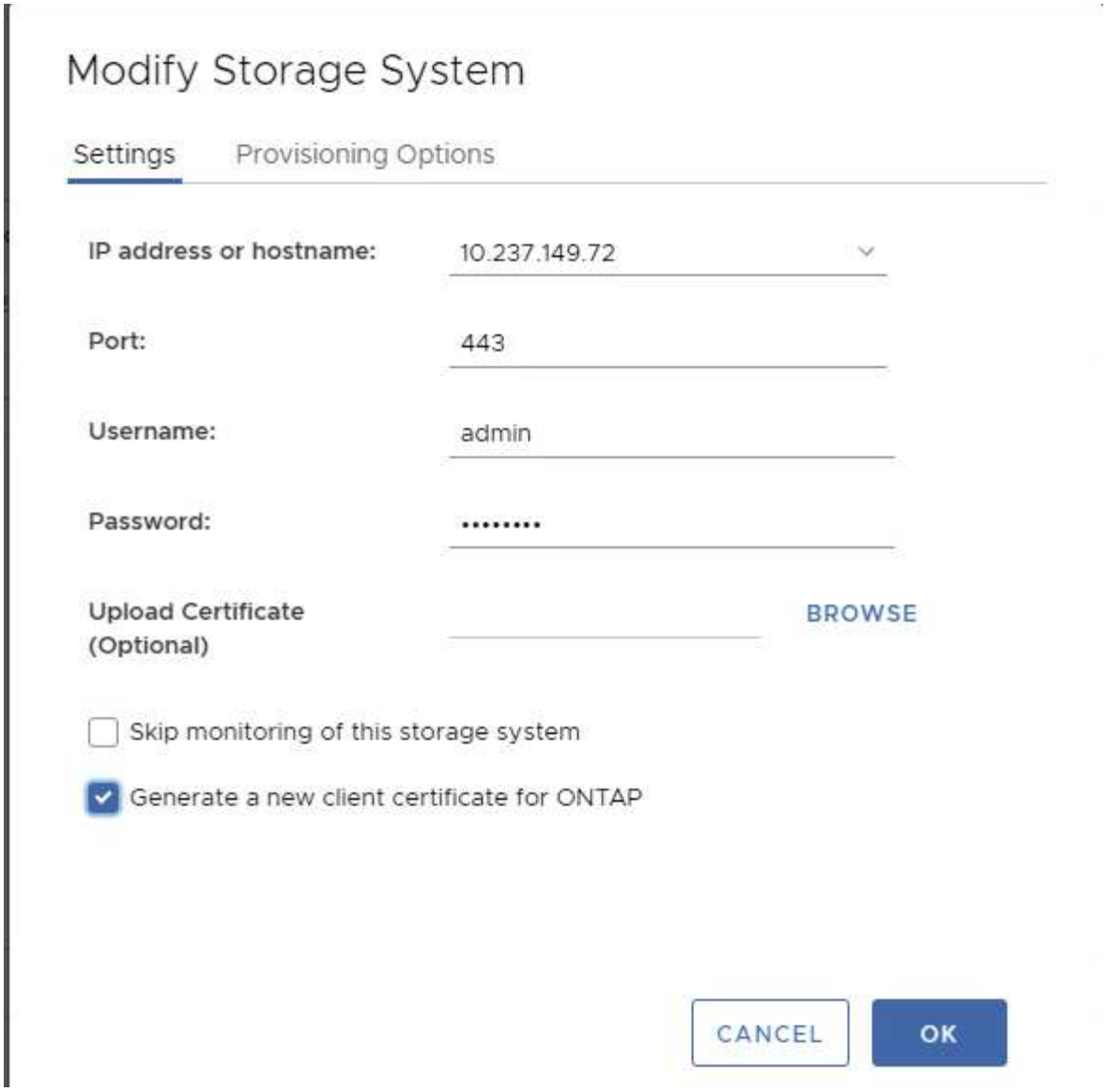

Add Storage System
Any communication between ONTAP tools plug-in and the storage system should be mutually authenticated.
vCenter server
Authorize Cluster Certificate
Host 10.234.85.142 has identified itself with a self-signed certificate.
Show certificate
Do you want to trust this certificate?
NO
YES
CANCEL
ADD



Modification du cluster

Lors de l’opération d’édition de cluster, il existe deux scénarios :

- Si le certificat ONTAP expire, l’utilisateur devra obtenir le nouveau certificat et le télécharger.
- Si le certificat OTV expire, l’utilisateur peut le régénérer en cochant la case.
  - *Générer un nouveau certificat client pour ONTAP.*


Modify Storage System
Settings
Provisioning Options
IP address or hostname:
10.237.149.72
Port:
443
Username:
admin
Password:
Upload Certificate (Optional)
BROWSE
Skip monitoring of this storage system
Generate a new client certificate for ONTAP
CANCEL
OK

Certificat HTTPS des outils ONTAP 9.13

Par défaut, les outils ONTAP utilisent un certificat auto-signé automatiquement créé lors de l'installation pour sécuriser l'accès HTTPS à l'interface utilisateur Web. Les outils ONTAP offrent les fonctionnalités suivantes :

1. Régénérer le certificat HTTPS

Lors de l'installation des outils ONTAP, un certificat d'autorité de certification HTTPS est installé et le certificat est stocké dans le magasin de clés. L'utilisateur a la possibilité de régénérer le certificat HTTPS via la console maint.

Les options ci-dessus sont accessibles dans *maint* console en accédant à *'Configuration de l'application' → 'régénérer les certificats'.*

Bannière de connexion Outils ONTAP 9.13

La bannière de connexion suivante s'affiche lorsque l'utilisateur saisit un nom d'utilisateur

dans l'invite de connexion. Notez que SSH est désactivé par défaut et n'autorise que les connexions uniques lorsqu'elles sont activées à partir de la console de la machine virtuelle.

```
WARNING: Unauthorized access to this system is forbidden and will be
prosecuted by law. By accessing this system, you agree that your actions
may be monitored if unauthorized usage is suspected.
```

Une fois que l'utilisateur a terminé sa connexion via le canal SSH, le texte suivant s'affiche :

```
Linux UnifiedVSC 5.10.0-21-amd64 #1 SMP Debian 5.10.162-1 (2023-01-21)
x86_64
```

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
```

```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

Délai d'inactivité pour les outils ONTAP 9.13

Pour empêcher tout accès non autorisé, un délai d'inactivité est défini, ce qui déconnecte automatiquement les utilisateurs inactifs pendant une certaine période pendant l'utilisation des ressources autorisées. Cela permet de garantir que seuls les utilisateurs autorisés peuvent accéder aux ressources et contribue à maintenir la sécurité.

- Par défaut, les sessions du client vSphere se ferment après 120 minutes d'inactivité, ce qui oblige l'utilisateur à se reconnecter pour reprendre à l'aide du client. Vous pouvez modifier la valeur du délai d'attente en modifiant le fichier webclient.properties. Vous pouvez configurer le délai d'expiration du client vSphere "Configurez la valeur du délai d'expiration du client vSphere"
- Les outils ONTAP ont un délai de déconnexion de session de l'interface de ligne de commande Web de 30 minutes.

Nombre maximal de requêtes simultanées par utilisateur (protection de la sécurité réseau/attaque dos) Outils ONTAP pour VMware vSphere 9.13

Par défaut, le nombre maximal de requêtes simultanées par utilisateur est de 48. L'utilisateur root des outils ONTAP peut modifier cette valeur en fonction des besoins de son environnement. **Cette valeur ne doit pas être définie sur une valeur très élevée car cela fournit un mécanisme contre les attaques par déni de service (DOS).**

Les utilisateurs peuvent modifier le nombre maximal de sessions simultanées et d'autres paramètres pris en

charge dans le fichier ***/opt/netapp/vscserver/etc/dofilterParams.json***.

Nous pouvons configurer le filtre en utilisant les paramètres suivants :

- ***delayMS***: Le délai en millisecondes donné à toutes les demandes au-delà de la limite de taux avant qu'elles ne soient prises en compte. Donnez -1 pour rejeter simplement la demande.
- ***étrangletMs***: Combien de temps pour attendre le sémaphore en mode asynchrone.
- ***maxRequestMS*** : durée d'exécution de cette requête.
- ***ipWhitelist***: Une liste d'adresses IP séparées par des virgules qui ne seront pas à débit limité. (Il peut s'agir d'adresses IP vCenter, ESXi et SRA)
- ***maxRequestsPerSec*** : nombre maximal de requêtes provenant d'une connexion par seconde.

Valeurs par défaut dans le fichier *dofilterParams*:

```
{"delayMs":"-1",
"throttleMs":"1800000",
"maxRequestMs":"300000",
"ipWhitelist":"10.224.58.52",
"maxRequestsPerSec":"48"}
```

Configuration du protocole NTP (Network Time Protocol) pour les outils ONTAP 9.13

Des problèmes de sécurité peuvent parfois se produire en raison de différences dans les configurations de l'heure du réseau. Il est important de s'assurer que tous les périphériques d'un réseau disposent de paramètres d'heure précis pour éviter de tels problèmes.

Appareil virtuel

Vous pouvez configurer le ou les serveurs NTP à partir de la console de maintenance de l'appliance virtuelle. Les utilisateurs peuvent ajouter les détails du serveur NTP sous *System Configuration* ⇒ *Add New NTP Server* option

Par défaut, le service NTP est ntpd. Il s'agit d'un service hérité qui ne fonctionne pas bien pour les machines virtuelles dans certains cas.

Debian

Sous Debian, l'utilisateur peut accéder au fichier /etc/ntp.conf pour obtenir des détails sur le serveur ntp.

Stratégies de mot de passe pour les outils ONTAP 9.13

Les utilisateurs qui déploient des outils ONTAP pour la première fois ou qui effectuent une mise à niveau vers la version 9.12 ou ultérieure devront suivre la stratégie de mot de passe robuste pour l'administrateur et les utilisateurs de base de données. Au cours du processus de déploiement, les nouveaux utilisateurs seront invités à entrer leurs mots de passe. Pour les utilisateurs de brownfield qui effectuent une mise à niveau vers la version

9.12 ou ultérieure, l’option de suivre la stratégie de mot de passe fort sera disponible dans la console de maintenance.

- Une fois que l’utilisateur se connecte à la console maint, les mots de passe sont vérifiés par rapport au jeu de règles complexes et s’il n’est pas suivi, l’utilisateur est invité à les réinitialiser.
- La validité par défaut du mot de passe est de 90 jours et après 75 jours, l’utilisateur commence à recevoir la notification de modification du mot de passe.
- Il est nécessaire de définir un nouveau mot de passe à chaque cycle, le système ne prendra pas le dernier mot de passe comme nouveau mot de passe.
- Chaque fois qu’un utilisateur se connecte à la console maint, il vérifie les stratégies de mot de passe comme les captures d’écran ci-dessous avant de charger le menu principal :

```
Maintenance Console : "NetApp ONTAP tools for VMware vSphere"

Discovered interfaces: eth0 (ENABLED)
validating password policies
```

- S’il n’est pas trouvé en suivant la stratégie de mot de passe ou sa configuration de mise à niveau à partir des outils ONTAP 9.11 ou antérieurs. L’utilisateur verra alors l’écran suivant pour réinitialiser le mot de passe :

```
Your Administrator and Database password is expired or does not match password policy:
----------
   1 ) Change 'administrator' user password
   2 ) Change database password

   x ) Exit

Enter your choice: _
```

- Si l’utilisateur tente de définir un mot de passe faible ou donne à nouveau le dernier mot de passe, l’erreur suivante s’affiche :

```
Changing password for administrator.

User: administrator
Enter new password:
Retype new password:

Password doesn't matches the password policy.
For security reasons, it is recommended to use a password that is of eight to thirty characters and
contains a minimum of one upper, one lower, one digit, and one special character.

Enter new password:
Retype new password:
Check if new decoder works ?
New decoder worked successfully
08/02/23 13:36:53 Your new password must be different

Error updating sra credential file


 Press ENTER to continue._
```

Mentions légales

Les mentions légales donnent accès aux déclarations de copyright, aux marques, aux brevets, etc.

Droits d'auteur

"https://www.netapp.com/company/legal/copyright/"

Marques déposées

NetApp, le logo NETAPP et les marques mentionnées sur la page des marques commerciales NetApp sont des marques commerciales de NetApp, Inc. Les autres noms de sociétés et de produits peuvent être des marques commerciales de leurs propriétaires respectifs.

"https://www.netapp.com/company/legal/trademarks/"

Brevets

Vous trouverez une liste actuelle des brevets appartenant à NetApp à l'adresse suivante :

https://www.netapp.com/pdf.html?item=/media/11887-patentspage.pdf

Politique de confidentialité

"https://www.netapp.com/company/legal/privacy-policy/"

Source ouverte

Les fichiers de notification fournissent des informations sur les droits d'auteur et les licences de tiers utilisés dans le logiciel NetApp.

ONTAP

"Avis pour ONTAP 9.16.1" "Avis pour ONTAP 9.16.0" "Avis pour ONTAP 9.15.1" "Avis pour ONTAP 9.15.0" "Avis pour ONTAP 9.14.1" "Avis pour ONTAP 9.14.0" "Avis pour ONTAP 9.13.1" "Notification relative à ONTAP 9.12.1" "Notification relative à ONTAP 9.12.0" "Notification relative à ONTAP 9.11.1" "Notification relative à ONTAP 9.10.1" "Avis pour ONTAP 9.10.0" "Notification relative à ONTAP 9.9.1" "Notification relative à ONTAP 9.8" "Avis pour ONTAP 9.7" "Avis pour ONTAP 9.6" "Avis pour ONTAP 9.5" "Avis pour ONTAP 9.4" "Avis pour ONTAP 9.3" "Avis pour ONTAP 9.2" "Avis pour ONTAP 9.1"

ONTAP Mediator pour les configurations IP MetroCluster

"9.9.1 Avis concernant le médiateur ONTAP pour les configurations IP MetroCluster" "9.8 Avis concernant le médiateur ONTAP pour les configurations IP MetroCluster" "9.7 Avis concernant le médiateur ONTAP pour les configurations IP MetroCluster"

Informations sur le copyright

Copyright © 2025 NetApp, Inc. Tous droits réservés. Imprimé aux États-Unis. Aucune partie de ce document protégé par copyright ne peut être reproduite sous quelque forme que ce soit ou selon quelque méthode que ce soit (graphique, électronique ou mécanique, notamment par photocopie, enregistrement ou stockage dans un système de récupération électronique) sans l'autorisation écrite préalable du détenteur du droit de copyright.

Les logiciels dérivés des éléments NetApp protégés par copyright sont soumis à la licence et à l'avis de non-responsabilité suivants :

CE LOGICIEL EST FOURNI PAR NETAPP « EN L'ÉTAT » ET SANS GARANTIES EXPRESSES OU TACITES, Y COMPRIS LES GARANTIES TACITES DE QUALITÉ MARCHANDE ET D'ADÉQUATION À UN USAGE PARTICULIER, QUI SONT EXCLUES PAR LES PRÉSENTES. EN AUCUN CAS NETAPP NE SERA TENU POUR RESPONSABLE DE DOMMAGES DIRECTS, INDIRECTS, ACCESSOIRES, PARTICULIERS OU EXEMPLAIRES (Y COMPRIS L'ACHAT DE BIENS ET DE SERVICES DE SUBSTITUTION, LA PERTE DE JOUISSANCE, DE DONNÉES OU DE PROFITS, OU L'INTERRUPTION D'ACTIVITÉ), QUELLES QU'EN SOIENT LA CAUSE ET LA DOCTRINE DE RESPONSABILITÉ, QU'IL S'AGISSE DE RESPONSABILITÉ CONTRACTUELLE, STRICTE OU DÉLICTUELLE (Y COMPRIS LA NÉGLIGENCE OU AUTRE) DÉCOULANT DE L'UTILISATION DE CE LOGICIEL, MÊME SI LA SOCIÉTÉ A ÉTÉ INFORMÉE DE LA POSSIBILITÉ DE TELS DOMMAGES.

NetApp se réserve le droit de modifier les produits décrits dans le présent document à tout moment et sans préavis. NetApp décline toute responsabilité découlant de l'utilisation des produits décrits dans le présent document, sauf accord explicite écrit de NetApp. L'utilisation ou l'achat de ce produit ne concède pas de licence dans le cadre de droits de brevet, de droits de marque commerciale ou de tout autre droit de propriété intellectuelle de NetApp.

Le produit décrit dans ce manuel peut être protégé par un ou plusieurs brevets américains, étrangers ou par une demande en attente.

LÉGENDE DE RESTRICTION DES DROITS : L'utilisation, la duplication ou la divulgation par le gouvernement sont sujettes aux restrictions énoncées dans le sous-paragraphe (b)(3) de la clause Rights in Technical Data-Noncommercial Items du DFARS 252.227-7013 (février 2014) et du FAR 52.227-19 (décembre 2007).

Les données contenues dans les présentes se rapportent à un produit et/ou service commercial (tel que défini par la clause FAR 2.101). Il s'agit de données propriétaires de NetApp, Inc. Toutes les données techniques et tous les logiciels fournis par NetApp en vertu du présent Accord sont à caractère commercial et ont été exclusivement développés à l'aide de fonds privés. Le gouvernement des États-Unis dispose d'une licence limitée irrévocable, non exclusive, non cessible, non transférable et mondiale. Cette licence lui permet d'utiliser uniquement les données relatives au contrat du gouvernement des États-Unis d'après lequel les données lui ont été fournies ou celles qui sont nécessaires à son exécution. Sauf dispositions contraires énoncées dans les présentes, l'utilisation, la divulgation, la reproduction, la modification, l'exécution, l'affichage des données sont interdits sans avoir obtenu le consentement écrit préalable de NetApp, Inc. Les droits de licences du Département de la Défense du gouvernement des États-Unis se limitent aux droits identifiés par la clause 252.227-7015(b) du DFARS (février 2014).

Informations sur les marques commerciales

NETAPP, le logo NETAPP et les marques citées sur le site http://www.netapp.com/TM sont des marques déposées ou des marques commerciales de NetApp, Inc. Les autres noms de marques et de produits sont des marques commerciales de leurs propriétaires respectifs.